

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04493 A2

(51) International Patent Classification⁷: **C07K 14/155**

(74) Agents: DOLLARD, Anne, S. et al.; Chiron Corporation,
Intellectual Property - R440, P.O. Box 8097, Emeryville,
CA 94662-8097 (US).

(21) International Application Number: **PCT/US01/21241**

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(22) International Filing Date: 5 July 2001 (05.07.2001)

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/610,313 5 July 2000 (05.07.2000) US

(71) Applicants (*for all designated States except US*): CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US). UNIVERSITY OF STELLENBOSCH [ZA/ZA]; P.O. Box 19063, 7505 Tygerberg (ZA).

Published:

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A2

WO 02/04493

(54) Title: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

(57) Abstract: The present invention relates to polynucleotides encoding immunogenic HIV type C polypeptides. Uses of the polynucleotides in applications including DNA immunization, generation of packaging cell lines, and production of HIV Type C proteins are also described.

**POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C POLYPEPTIDES,
POLYPEPTIDES AND USES THEREOF**

TECHNICAL FIELD

5 Polynucleotides encoding antigenic Type C HIV polypeptides (e.g., Gag, pol, vif, vpr, tat, rev, vpu, env, and nef) are described, as are uses of these polynucleotides and polypeptide products in immunogenic compositions. Also described are polynucleotide sequences from South African variants of HIV Type C.

10 **BACKGROUND OF THE INVENTION**

Acquired immune deficiency syndrome (AIDS) is recognized as one of the greatest health threats facing modern medicine. There is, as yet, no cure for this disease.

In 1983-1984, three groups independently identified the suspected etiological agent of AIDS. See, e.g., Barre-Sinoussi et al. (1983) Science 220:868-871; Montagnier et al., in Human 15 T-Cell Leukemia Viruses (Gallo, Essex & Gross, eds., 1984); Vilmer et al. (1984) The Lancet 1:753; Popovic et al. (1984) Science 224:497-500; Levy et al. (1984) Science 225:840-842. These isolates were variously called lymphadenopathy-associated virus (LAV), human T-cell lymphotropic virus type III (HTLV-III), or AIDS-associated retrovirus (ARV). All of these isolates are strains of the same virus, and were later collectively named Human 20 Immunodeficiency-Virus (HIV). With the isolation of a related AIDS-causing virus, the strains originally called HIV are now termed HIV-1 and the related virus is called HIV-2 See, e.g., Guyader et al. (1987) Nature 326:662-669; Brun-Vezinet et al. (1986) Science 233:343-346; Clavel et al. (1986) Nature 324:691-695.

A great deal of information has been gathered about the HIV virus, however, to date 25 an effective vaccine has not been identified. Several targets for vaccine development have been examined including the *env* and *Gag* gene products encoded by HIV. *Gag* gene products include, but are not limited to, *Gag*-polymerase and *Gag*-protease. *Env* gene products include, but are not limited to, monomeric gp120 polypeptides, oligomeric gp140 polypeptides and gp160 polypeptides.

30 Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.* 72(2):1497-1503, 1998) described an increased

immune response elicited by DNA vaccination employing a synthetic gp120 sequence with modified codon usage. Schneider, et al., (*J Virol.* 71(7):4892-4903, 1997) discuss inactivation of inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences.

5 The *Gag* proteins of HIV-1 are necessary for the assembly of virus-like particles. HIV-1 *Gag* proteins are involved in many stages of the life cycle of the virus including, assembly, virion maturation after particle release, and early post-entry steps in virus replication. The roles of HIV-1 *Gag* proteins are numerous and complex (Freed, E.O., *Virology* 251:1-15, 1998).

10 Wolf, et al., (PCT International Application, WO 96/30523, published 3 October 1996; European Patent Application, Publication No. 0 449 116 A1, published 2 October 1991) have described the use of altered pr55 *Gag* of HIV-1 to act as a non-infectious retroviral-like particulate carrier, in particular, for the presentation of immunologically important epitopes. Wang, et al., (*Virology* 200:524-534, 1994) describe a system to study 15 assembly of HIV Gag- β -galactosidase fusion proteins into virions. They describe the construction of sequences encoding HIV Gag- β -galactosidase fusion proteins, the expression of such sequences in the presence of HIV Gag proteins, and assembly of these proteins into virus particles.

20 Shiver, et al., (PCT International Application, WO 98/34640, published 13 August 1998) described altering HIV-1 (CAM1) *Gag* coding sequences to produce synthetic DNA molecules encoding HIV *Gag* and modifications of HIV *Gag*. The codons of the synthetic molecules were codons preferred by a projected host cell.

25 Recently, use of HIV Env polypeptides in immunogenic compositions has been described. (see, U.S. Patent No. 5,846,546 to Hurwitz et al., issued December 8, 1998, describing immunogenic compositions comprising a mixture of at least four different recombinant virus that each express a different HIV env variant; and U.S. Patent No. 5,840,313 to Vahlne et al., issued November 24, 1998, describing peptides which correspond to epitopes of the HIV-1 gp120 protein). In addition, U.S. Patent No. 5,876,731 to Sia et al., issued March 2, 1999 describes candidate vaccines against HIV comprising an amino acid 30 sequence of a T-cell epitope of Gag linked directly to an amino acid sequence of a B-cell epitope of the V3 loop protein of an HIV-1 isolate containing the sequence GPGR. There remains a need for antigenic HIV polypeptides, particularly Type C isolates.

SUMMARY OF THE INVENTION

Described herein are novel Type C HIV sequences, for example, 8_5_TV1_C.ZA, 8_2_TV1_C.ZA and 12-5_1_TV2_C.ZA, polypeptides encoded by these novel sequences, and synthetic expression cassettes generated from these and other Type C HIV sequences.

In certain embodiments, the present invention relates synthetic expression cassettes encoding HIV Type C polypeptides, including Env, Gag, Pol, Prot, Vpr, Vpu, Vif, Nef, Tat, Rev and/or fragments thereof. In addition, the present invention also relates to improved expression of HIV Type C polypeptides and production of virus-like particles. Synthetic expression cassettes encoding the HIV polypeptides (e.g., Gag-, pol-, protease (prot)-, reverse transcriptase, integrase, RNaseH, Tat, Rev, Nef, Vpr, Vpu, Vif and/or Env- containing polypeptides) are described, as are uses of the expression cassettes.

Thus, one aspect of the present invention relates to expression cassettes and polynucleotides contained therein. The expression cassettes typically include an HIV-polypeptide encoding sequence inserted into an expression vector backbone. In one embodiment, an expression cassette comprises a polynucleotide sequence encoding one or more *Pol*-containing polypeptides, wherein the polynucleotide sequence comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and more preferably about 98% sequence (and any integers between these values) identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Pol*-containing polypeptides include, but are not limited to, those shown in SEQ ID NO:30, SEQ ID NO:31; SEQ ID NO:32; SEQ ID NO:62; SEQ ID NO:103; SEQ ID NO:58; SEQ ID NO:60; SEQ ID NO:64; SEQ ID NO:66; SEQ ID NO:68; SEQ ID NO:70; SEQ ID NO:76; and SEQ ID NO:78.

The polynucleotides encoding the HIV polypeptides of the present invention may also include sequences encoding additional polypeptides. Such additional polynucleotides encoding polypeptides may include, for example, coding sequences for other viral proteins (e.g., hepatitis B or C or other HIV proteins, such as, polynucleotide sequences encoding an HIV *Gag* polypeptide, polynucleotide sequences encoding an HIV *Env* polypeptide and/or polynucleotides encoding one or more of vif, vpr, tat, rev, vpu and nef); cytokines or other transgenes. In one embodiment, the sequence encoding the HIV *Pol* polypeptide(s) can be modified by deletions of coding regions corresponding to reverse transcriptase and integrase.

Such deletions in the polymerase polypeptide can also be made such that the polynucleotide sequence preserves T-helper cell and CTL epitopes. Other antigens of interest may be inserted into the polymerase as well.

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Gag*-containing polypeptide, wherein the polynucleotide sequence encoding the *Gag* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Gag*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 844-903 of Figure 1 (a *Gag* major homology region) (SEQ ID NO:1); nucleotides 841-900 of Figure 2 (a *Gag* major homology region) (SEQ ID NO:2); Figure 24 (SEQ ID NO:53, a *Gag* major homology region); the sequence presented as Figure 1 (SEQ ID NO:3); the sequence presented as Figure 22 (SEQ ID NO:51); the sequence presented as Figure 70 (SEQ ID NO:99); and the sequence presented as Figure 2 (SEQ ID NO:4). As noted above, the polynucleotides encoding the *Gag*-containing polypeptides of the present invention may also include sequences encoding additional polypeptides.

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Env*-containing polypeptide, wherein the polynucleotide sequence encoding the *Env* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Env*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 1213-1353 of Figure 3 (SEQ ID NO:5) (encoding an Env common region); the sequence presented as Figure 17 (SEQ ID NO:46) (encoding a 97 nucleotide long Env common region); SEQ ID NO:47 (encoding a 144 nucleotide long Env common region); nucleotides 82-1512 of Figure 3 (SEQ ID NO:6) (encoding a gp120 polypeptide); nucleotides 82-2025 of Figure 3 (SEQ ID NO:7) (encoding a gp140 polypeptide); nucleotides 82-2547 of Figure 3 (SEQ ID NO:8) (encoding a gp160 polypeptide); SEQ ID NO:49 (encoding a gp160 polypeptide); nucleotides 1-2547 of Figure 3 (SEQ ID NO:9) (encoding a gp160 polypeptide with signal sequence); nucleotides 1513-2547 of Figure 3 (SEQ ID NO:10) (encoding a gp41 polypeptide); nucleotides 1210-1353 of

Figure 4 (SEQ ID NO:11) (encoding an Env common region); nucleotides 73-1509 of Figure 4 (SEQ ID NO:12) (encoding a gp120 polypeptide); nucleotides 73-2022 of Figure 4 (SEQ ID NO:13) (encoding a gp140 polypeptide); nucleotides 73-2565 of Figure 4 (SEQ ID NO:14) (encoding a gp160 polypeptide); nucleotides 1-2565 of Figure 4 (SEQ ID NO:15) 5 (encoding a gp160 polypeptide with signal sequence); the sequence presented as Figure 20 (SEQ ID NO:49) (encoding a gp160 polypeptide); the sequence presented as Figure 68 (SEQ ID NO:97) (encoding a gp160 polypeptide); nucleotides 1510-2565 of Figure 4 (SEQ ID NO:16) (encoding a gp41 polypeptide); nucleotides 7 to 1464 of Figure 90 (SEQ ID NO:119) (encoding a gp120 polypeptide with modified wild type signal sequence); nucleotides 7 to 10 1977 of Figure 91 (SEQ ID NO:120) (encoding a gp140 polypeptide including signal sequence modified from wild-type 8_2_TV1_C.ZA (e.g., "modified wild type leader sequence")); nucleotides 7 to 1977 of Figure 92 (SEQ ID NO:121) (encoding a gp140 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2388 15 of Figure 93 (SEQ ID NO:122) (encoding a gp160 polypeptide with modified wild type signal sequence); nucleotides 7 to 2520 of Figure 94 (SEQ ID NO:123) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2520 of Figure 95 (SEQ ID NO:124) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 13 to 2604 of Figure 96 (SEQ ID NO:125) (encoding a gp160 polypeptide with TPA1 signal sequence); nucleotides 7 to 2607 of Figure 20 97 (SEQ ID NO:126) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 1 to 2049 of Figure 100 (SEQ ID NO:131) (encoding a gp140 polypeptide with TPA1 signal sequence); nucleotides 7 to 1607 of Figure 98 (SEQ ID NO:126) (encoding a gp160 polypeptide with wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2064 of SEQ ID NO:132 (encoding a gp140 polypeptide with 25 modified wild-type 8_2_TV1_C.ZA leader sequence); and nucleotides 7 to 2064 of SEQ ID NO:133 (encoding a gp140 polypeptide with wild-type 8_2_TV1_C.ZA leader sequence).

In certain embodiments, the Env-encoding sequences will contain further modifications, for instance mutation of the cleavage site to prevent the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide (SEQ ID NO:121 and SEQ ID NO:124) or deletion of variable regions V1 and/or V2 (SEQ ID NO:119; SEQ ID NO:120; SEQ ID NO:121; SEQ ID NO:122; SEQ ID NO:123; and SEQ ID NO:124).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Nef*-containing polypeptide, wherein the polynucleotide sequence encoding the *Nef* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Nef*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 26 (SEQ ID NO:55); the sequence presented in Figure 72 (SEQ ID NO:101); the sequence presented in Figure 28 (SEQ ID NO:57); the sequence presented in Figure 67 (SEQ ID NO:96); the sequence presented in Figure 103 (SEQ ID NO:134); and the sequence presented in Figure 104 (SEQ ID NO:135).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Rev*-containing polypeptide, wherein the polynucleotide sequence encoding the *Rev* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Rev*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 43 (SEQ ID NO:72); the sequence presented in Figure 76 (SEQ ID NO:105); the sequence presented in Figure 45 (SEQ ID NO:74); the sequence presented in Figure 78 (SEQ ID NO:107); and the sequence presented in Figure 62 (SEQ ID NO:91).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Tat*-containing polypeptide, wherein the polynucleotide sequence encoding the *Tat* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Tat*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 51 (SEQ ID NO:80); the sequence presented in Figure 80 (SEQ ID NO:109); the sequence presented in Figure 52 (SEQ ID NO:81); the sequence presented in Figure 54 (SEQ ID NO:83); and the sequence presented in Figure 82 (SEQ ID NO:111).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Vif*-containing polypeptide, wherein the polynucleotide sequence encoding the *Vif* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Vif*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 56 (SEQ ID NO:85); and the sequence presented in Figure 84 (SEQ ID NO:113).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Vpr*-containing polypeptide, wherein the polynucleotide sequence encoding the *Vpr* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Vpr*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 58 (SEQ ID NO:87); and the sequence presented in Figure 86 (SEQ ID NO:115).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Vpu*-containing polypeptide, wherein the polynucleotide sequence encoding the *Vpu* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Vpu*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 60 (SEQ ID NO:89); and the sequence presented in Figure 88 (SEQ ID NO:117).

Further embodiments of the present invention include purified polynucleotides of any of the sequences described herein. Exemplary polynucleotide sequences encoding *Gag*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 844-903 of Figure 1 (SEQ ID NO:1) (a *Gag* major homology region); nucleotides 841-900 of Figure 2 (SEQ ID NO:2) (a *Gag* major homology region); the sequence presented as Figure 1 (SEQ ID NO:3); the sequence presented as Figure 2 (SEQ ID NO:4); the sequence presented as Figure 22 (SEQ ID NO:51); the sequence presented as Figure 70 (SEQ

ID NO:99); and the sequence presented as Figure 24 (SEQ ID NO:53) (a Gag major homology region).

Exemplary polynucleotide sequences encoding *Env*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 1213-1353 of Figure 3 (SEQ ID NO:5) (encoding an Env common region); the sequence presented as Figure 17 (SEQ ID NO:46) (encoding a 97 nucleotide long Env common region); SEQ ID NO:47 (encoding a 144 nucleotide long Env common region); nucleotides 82-1512 of Figure 3 (SEQ ID NO:6) (encoding a gp120 polypeptide); nucleotides 82-2025 of Figure 3 (SEQ ID NO:7) (encoding a gp140 polypeptide); nucleotides 82-2547 of Figure 3 (SEQ ID NO:8) (encoding a gp160 polypeptide); SEQ ID NO:49 (encoding a gp160 polypeptide); nucleotides 1-2547 of Figure 3 (SEQ ID NO:9) (encoding a gp160 polypeptide with signal sequence); nucleotides 1513-2547 of Figure 3 (SEQ ID NO:10) (encoding a gp41 polypeptide); nucleotides 1210-1353 of Figure 4 (SEQ ID NO:11) (encoding an Env common region); nucleotides 73-1509 of Figure 4 (SEQ ID NO:12) (encoding a gp120 polypeptide); nucleotides 73-2022 of Figure 4 (SEQ ID NO:13) (encoding a gp140 polypeptide); nucleotides 73-2565 of Figure 4 (SEQ ID NO:14) (encoding a gp160 polypeptide); nucleotides 1-2565 of Figure 4 (SEQ ID NO:15) (encoding a gp160 polypeptide with signal sequence); the sequence presented as Figure 20 (SEQ ID NO:49) (encoding a gp160 polypeptide); the sequence presented as Figure 68 (SEQ ID NO:97) (encoding a gp160 polypeptide); nucleotides 1510-2565 of Figure 4 (SEQ ID NO:16) (encoding a gp41 polypeptide); nucleotides 7 to 1464 of Figure 90 (SEQ ID NO:119) (encoding a gp120 polypeptide with modified wild type signal sequence); nucleotides 7 to 1977 of Figure 91 (SEQ ID NO:120) (encoding a gp140 polypeptide including signal sequence modified from wild-type 8_2_TV1_C.ZA (e.g., "modified wild type leader sequence")); nucleotides 7 to 1977 of Figure 92 (SEQ ID NO:121) (encoding a gp140 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2388 of Figure 93 (SEQ ID NO:122) (encoding a gp160 polypeptide with modified wild type signal sequence); nucleotides 7 to 2520 of Figure 94 (SEQ ID NO:123) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2520 of Figure 95 (SEQ ID NO:124) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 13 to 2604 of Figure 96 (SEQ ID NO:125) (encoding a gp160 polypeptide with TPA1 signal sequence); nucleotides 7 to 2607 of Figure 97 (SEQ ID NO:126) (encoding a gp160 polypeptide with modified wild type

8_2_TV1_C.ZA signal sequence); nucleotides 1 to 2049 of Figure 100 (SEQ ID NO:131) (encoding a gp140 polypeptide with TPA1 signal sequence); nucleotides 7 to 1607 of Figure 98 (SEQ ID NO:126) (encoding a gp160 polypeptide with wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2064 of SEQ ID NO:132 (encoding a gp140 polypeptide with modified wild-type 8_2_TV1_C.ZA leader sequence); and nucleotides 7 to 2064 of SEQ ID NO:133 (encoding a gp140 polypeptide with wild-type 8_2_TV1_C.ZA leader sequence).

Exemplary purified polynucleotides encoding additional HIV polynucleotides include: Pol-encoding polynucleotides (*e.g.*, SEQ ID NO:30, SEQ ID NO:31; SEQ ID NO:32; SEQ ID NO:62; SEQ ID NO:103; SEQ ID NO:58; SEQ ID NO:60; SEQ ID NO:64; SEQ ID NO:66; SEQ ID NO:68; SEQ ID NO:70; SEQ ID NO:76; and SEQ ID NO:78); Nef-encoding polynucleotides (*e.g.*, SEQ ID NO:55; SEQ ID NO:101; SEQ ID NO:57; SEQ ID NO:96); Rev-encoding polynucleotides (*e.g.*, SEQ ID NO:72; SEQ ID NO:105; SEQ ID NO:74); SEQ ID NO:107; SEQ ID NO:91); Tat-encoding polynucleotides (*e.g.*, SEQ ID NO:80; SEQ ID NO:109; SEQ ID NO:81; SEQ ID NO:83; SEQ ID NO:111); Vif-encoding polynucleotides (*e.g.*, SEQ ID NO:85; SEQ ID NO:113); and Vpr-encoding polynucleotides (*e.g.*, SEQ ID NO:87; SEQ ID NO:115); Vpu-encoding polynucleotides (*e.g.*, SEQ ID NO:89; SEQ ID NO:117).

In other embodiments, the present invention relates to native HIV polypeptide-encoding sequences obtained from novel Type C strains; fragments of these native sequences; expression cassettes containing these wild-type sequences; and uses of these sequences, fragments and expression cassettes. Exemplary full length sequences are shown in SEQ ID NO:33 and SEQ ID NO:45. Exemplary fragments coding for various HIV gene products include: the sequence presented in Figure 19 (SEQ ID NO:48) (an Env-encoding sequence); the sequence presented in Figure 69 (SEQ ID NO:98) (an Env-encoding sequence); the sequence presented in Figure 21 (SEQ ID NO:50) (a gp160 polypeptide); the sequence presented in Figure 23 (SEQ ID NO:52) (a Gag polypeptide); the sequence presented in Figure 71 (SEQ ID NO:100) (a Gag polypeptide); the sequence presented in Figure 25 (SEQ ID NO:54) (a Gag polypeptide); the sequence presented in Figure 27 (SEQ ID NO:56) (a Nef polypeptide); the sequence presented in Figure 73 (SEQ ID NO:102) (a Nef polypeptide); the sequence presented in Figure 30 (SEQ ID NO:59) (a p15RNaseH polypeptide); the sequence presented in Figure 32 (SEQ ID NO:61) (a p31Integrase polypeptide); the sequence presented in Figure 34 (SEQ ID NO:63) (a Pol polypeptide); the sequence presented in Figure 75 (SEQ

ID NO:104) (a Pol polypeptide); the sequence presented in Figure 36 (SEQ ID NO:65) (a Prot polypeptide); the sequence presented in Figure 38 (SEQ ID NO:67) (a inactivated Prot polypeptide); the sequence presented in Figure 40 (SEQ ID NO:69) (an inactivated Prot and RT polypeptide); the sequence presented in Figure 42 (SEQ ID NO:71) (a Prot and RT polypeptide); the sequence presented in Figure 44 (SEQ ID NO:73) (a Rev polypeptide); the sequence presented in Figure 77 (SEQ ID NO:106) (a Rev polypeptide); the sequence presented in Figure 46 (SEQ ID NO:75) (a Rev polypeptide); the sequence presented in Figure 79 (SEQ ID NO:108) (a Rev polypeptide); the sequence presented in Figure 48 (SEQ ID NO:77) (an RT polypeptide); the sequence presented in Figure 50 (SEQ ID NO:79) (a mutated RT polypeptide); the sequence presented in Figure 53 (SEQ ID NO:82) (a Tat polypeptide); the sequence presented in Figure 81 (SEQ ID NO:110) (a Tat polypeptide); the sequence presented in Figure 55 (SEQ ID NO:84) (a Tat polypeptide); the sequence presented in Figure 83 (SEQ ID NO:112) (a Tat polypeptide); the sequence presented in Figure 57 (SEQ ID NO:86) (a Vif polypeptide); the sequence presented in Figure 85 (SEQ ID NO:114) (a Vif polypeptide); the sequence presented in Figure 59 (SEQ ID NO:88) (a Vpr polypeptide); the sequence presented in Figure 82 (SEQ ID NO:116) (a Vpr polypeptide); the sequence presented in Figure 61 (SEQ ID NO:90) (a Vpu polypeptide); the sequence presented in Figure 89 (SEQ ID NO:118) (a Vpu polypeptide); the sequence presented in Figure 63 (SEQ ID NO:92) (a Rev polypeptide); and the sequence presented in Figure 66 (SEQ ID NO:95) (a Tat polypeptide).

The native and synthetic polynucleotide sequences encoding the HIV polypeptides of the present invention typically have at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught herein. Further, in certain embodiments, the polynucleotide sequences encoding the HIV polypeptides of the invention will exhibit 100% sequence identity to the sequences taught herein.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

The present invention further includes recombinant expression systems for use in selected host cells, wherein the recombinant expression systems employ one or more of the polynucleotides and expression cassettes of the present invention. In such systems, the polynucleotide sequences are operably linked to control elements compatible with expression

in the selected host cell. Numerous expression control elements are known to those in the art, including, but not limited to, the following: transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences. Exemplary transcription promoters include, but are not limited to those derived from CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

In another aspect the invention includes cells comprising one or more of the expression cassettes of the present invention where the polynucleotide sequences are operably linked to control elements compatible with expression in the selected cell. In one embodiment such cells are mammalian cells. Exemplary mammalian cells include, but are not limited to, BHK, VERO, HT1080, 293, RD, COS-7, and CHO cells. Other cells, cell types, tissue types, etc., that may be useful in the practice of the present invention include, but are not limited to, those obtained from the following: insects (e.g., *Trichoplusia ni* (Tn5) and Sf9), bacteria, yeast, plants, antigen presenting cells (e.g., macrophage, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), primary cells, immortalized cells, tumor-derived cells.

In a further aspect, the present invention includes compositions for generating an immunological response, where the composition typically comprises at least one of the expression cassettes of the present invention and may, for example, contain combinations of expression cassettes (such as one or more expression cassettes carrying a Pol-polypeptide-encoding polynucleotide, one or more expression cassettes carrying a Gag-polypeptide-encoding polynucleotide, one or more expression cassettes carrying accessory polypeptide-encoding polynucleotides (e.g., native or synthetic vpu, vpr, nef, vif, tat, rev), and/or one or more expression cassettes carrying an Env-polypeptide-encoding polynucleotide). Such compositions may further contain an adjuvant or adjuvants. The compositions may also contain one or more Type C HIV polypeptides. The Type C HIV polypeptides may correspond to the polypeptides encoded by the expression cassette(s) in the composition, or may be different from those encoded by the expression cassettes. An example of the polynucleotide in the expression cassette encoding the same polypeptide as is being provided in the composition is as follows: the polynucleotide in the expression cassette encodes the Gag-polypeptide of Figure 1 (SEQ ID NO:3), and the polypeptide (SEQ ID NO:17) is the polypeptide encoded by the sequence shown in Figure 1. An example of the polynucleotide in

the expression cassette encoding a different polypeptide as is being provided in the composition is as follows: an expression cassette having a polynucleotide encoding a Gag-polymerase polypeptide, and the polypeptide provided in the composition may be a Gag and/or Gag-protease polypeptide. In compositions containing both expression cassettes (or 5 polynucleotides of the present invention) and polypeptides, various expression cassettes of the present invention can be mixed and/or matched with various Type C HIV polypeptides described herein.

In another aspect the present invention includes methods of immunization of a subject. In the method any of the above described compositions are into the subject under 10 conditions that are compatible with expression of the expression cassette(s) in the subject. In one embodiment, the expression cassettes (or polynucleotides of the present invention) can be introduced using a gene delivery vector. The gene delivery vector can, for example, be a non-viral vector or a viral vector. Exemplary viral vectors include, but are not limited to Sindbis-virus derived vectors, retroviral vectors, and lentiviral vectors. Compositions useful 15 for generating an immunological response can also be delivered using a particulate carrier. Further, such compositions can be coated on, for example, gold or tungsten particles and the coated particles delivered to the subject using, for example, a gene gun. The compositions can also be formulated as liposomes. In one embodiment of this method, the subject is a mammal and can, for example, be a human.

20 In a further aspect, the invention includes methods of generating an immune response in a subject. Any of the expression cassettes described herein can be expressed in a suitable cell to provide for the expression of the Type C HIV polypeptides encoded by the polynucleotides of the present invention. The polypeptide(s) are then isolated (e.g., substantially purified) and administered to the subject in an amount sufficient to elicit an 25 immune response. In certain embodiments, the methods comprise administration of one or more of the expression cassettes or polynucleotides of the present invention, using any of the gene delivery techniques described herein. In other embodiments, the methods comprise co-administration of one or more of the expression cassettes or polynucleotides of the present invention and one or more polypeptides, wherein the polypeptides can be expressed from 30 these polynucleotides or can be other subtype C HIV polypeptides. In other embodiments, the methods comprise co-administration of multiple expression cassettes or polynucleotides of the present invention. In still further embodiments, the methods comprise co-

administration of multiple polypeptides, for example polypeptides expressed from the polynucleotides of the present invention and/or other subtype C HIV polypeptides.

The invention further includes methods of generating an immune response in a subject, where cells of a subject are transfected with any of the above-described expression 5 cassettes or polynucleotides of the present invention, under conditions that permit the expression of a selected polynucleotide and production of a polypeptide of interest (e.g., encoded by any expression cassette of the present invention). By this method an immunological response to the polypeptide is elicited in the subject. Transfection of the cells may be performed *ex vivo* and the transfected cells are reintroduced into the subject.

10 Alternately, or in addition, the cells may be transfected *in vivo* in the subject. The immune response may be humoral and/or cell-mediated (cellular). In a further embodiment, this method may also include administration of an Type C HIV polypeptides before, concurrently with, and/or after introduction of the expression cassette into the subject.

15 These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 (SEQ ID NO:3) shows the nucleotide sequence of a polynucleotide encoding a synthetic Gag polypeptide. The nucleotide sequence shown was obtained by modifying 20 type C strain AF110965 and include further modifications of INS.

Figure 2 (SEQ ID NO: 4) shows the nucleotide sequence of a polynucleotide encoding a synthetic Gag polypeptide. The nucleotide sequence shown was obtained by modifying type C strain AF110967 and include further modifications of INS.

Figure 3 (SEQ ID NO:9) shows the nucleotide sequence of a polynucleotide encoding 25 a synthetic Env polypeptide. The nucleotide sequence depicts gp160 (including a signal peptide) and was obtained by modifying type C strain AF110968. The arrows indicate the positions of various regions of the polynucleotide, including the sequence encoding a signal peptide (nucleotides 1-81) (SEQ ID NO:18), a gp120 polypeptide (nucleotides 82-1512) (SEQ ID NO:6), a gp41 polypeptide (nucleotides 1513-2547) (SEQ ID NO:10), a gp140 30 polypeptide (nucleotides 82-2025) (SEQ ID NO:7) and a gp160 polypeptide (nucleotides 82-2547) (SEQ ID NO:8). The codons encoding the signal peptide are modified (as described herein) from the native HIV-1 signal sequence.

Figure 4 (SEQ ID NO:15) shows the nucleotide sequence of a polynucleotide encoding a synthetic Env polypeptide. The nucleotide sequence depicts gp160 (including a signal peptide) and was obtained by modifying type C strain AF110975. The arrows indicate the positions of various regions of the polynucleotide, including the sequence encoding a signal peptide (nucleotides 1-72) (SEQ ID NO:19), a gp120 polypeptide (nucleotides 73-1509) (SEQ ID NO:12), a gp41 polypeptide (nucleotides 1510-2565) (SEQ ID NO:16), a gp140 polypeptide (nucleotides 73-2022) (SEQ ID NO:13), and a gp160 polypeptide (nucleotides 73-2565) (SEQ ID NO:14). The codons encoding the signal peptide are modified (as described herein) from the native HIV-1 signal sequence.

Figure 5 shows the location of some remaining INS in synthetic Gag sequences derived from AF110965. The changes made to these sequences are boxed in the Figures. The top line depicts a codon modified sequence of Gag polypeptides from the indicated strains (SEQ ID NO:20). The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to remove further INS and correspond to the sequence depicted in Figure 1 (SEQ ID NO:3).

Figure 6 shows the location of some remaining INS in synthetic Gag sequences derived from AF110967. The changes made to these sequences are boxed in the Figures. The top line depicts a modified sequence of Gag polypeptides from the indicated strains (SEQ ID NO:21). The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to remove further INS and correspond to the sequence depicted in Figure 2 (SEQ ID NO:4).

Figure 7 is a schematic depicting the selected domains in the *Pol* region of HIV.

Figure 8 (SEQ ID NO:30) depicts the nucleotide sequence of the synthetic construct designated PR975(+). "(+)" indicates that the reverse transcriptase is functional. This construct includes sequence from p2 (nucleotides 16 to 54 of SEQ ID NO:30); p7 (nucleotides 55 to 219 of SEQ ID NO:30); p1/p6 (nucleotides 220-375 of SEQ ID NO:30); prot (nucleotides 376 to 672 of SEQ ID NO:30), reverse transcriptase (nucleotides 673 to 2352 of SEQ ID NO:30); and 6 amino acids of integrase shown in Figure 7 (nucleotides 2353 to 2370 of SEQ ID NO:30). In addition, the construct contains a multiple cloning site (MCS, nucleotides 2425 to 2463 of SEQ ID NO:30) for insertion of a transgene and a YMDD epitope cassette (nucleotides 2371 to 2424 of SEQ ID NO:30).

Figure 9 (SEQ ID NO:31) depicts the nucleotide sequence of the synthetic construct designated PR975YM. As illustrated in Figure 7, the RT region includes a mutation in the catalytic center (mut. cat. center). "YM" refers to constructs in which the nucleotides encode the amino acids AP instead of YMDD in this region. Reverse transcriptase is not functional in this construct. This construct includes sequence from the p2 (nucleotides 16 to 54 of SEQ ID NO:31); p7 (nucleotides 55 to 219 of SEQ ID NO:31); p1/p6 (nucleotides 220 to 375 of SEQ ID NO:31); prot (nucleotides 376 to 672 of SEQ ID NO:31); and reverse transcriptase (nucleotides 673 to 2346 of SEQ ID NO:31) shown in Figure 7, although the reverse transcriptase protein is not functional. In addition, the construct contains a multiple cloning site (MCS, nucleotides 2419 to 2457 of SEQ ID NO:31) for insertion of a transgene and a YMDD epitope cassette (nucleotides 2365 to 2418 of SEQ ID NO:31).

Figure 10 (SEQ ID NO:32) depicts the nucleotide sequence of the synthetic construct designated PR975YMWM. "YM" refers to constructs in which the nucleotides encode the amino acids AP instead of YMDD in this region. "WM" refers to constructs in which the nucleotides encode amino acids PI instead of WMGY in this region. This construct includes sequence from the p2 (nucleotides 16 to 54 of SEQ ID NO:32); p7 (nucleotides 55 to 219 of SEQ ID NO:32); p1/p6 (nucleotides 220 to 375 of SEQ ID NO:32); prot (nucleotides 376 to 672 of SEQ ID NO:32); and reverse transcriptase (nucleotides 673 to 2340 of SEQ ID NO:32) shown in Figure 7, although the reverse transcriptase protein is not functional. In addition, the construct contains a multiple cloning site (MCS, nucleotides 2413 to 2451 of SEQ ID NO:32) for insertion of a transgene and a YMDD epitope cassette (nucleotides 2359 to 2412 of SEQ ID NO:32).

Figure 11 (SEQ ID NO:33) depicts the nucleotide sequence of 8_5_TV1_C.ZA. Various regions are shown in Table A.

Figure 12 (SEQ ID NO:34) depicts the wild type nucleotide sequence of AF110975 Pol from p2gag until p7gag.

Figure 13 (SEQ ID NO:35) depicts the wild type nucleotide sequence of AF110975 Pol from p1 through the first 6 amino acids of the integrase protein.

Figure 14 (SEQ ID NO:36) depicts the nucleotide sequence of a cassette encoding Ile178 through Serine 191 of reverse transcriptase.

Figure 15 (SEQ ID NO:37) shows amino acid sequence which includes an epitope in the region of the catalytic center of the reverse transcriptase protein.

Figure 16 (SEQ ID NO:45) depicts the nucleotide sequence of 12-5_1_TV2_C.ZA.

Figure 17 (SEQ ID NO:46) depicts the nucleotide sequence of a synthetic Env-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence corresponds to a short (97 base pair) common region.

5 Figure 18 (SEQ ID NO:47) depicts the nucleotide sequence of a synthetic Env-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence corresponds to a common region in Env.

Figure 19 (SEQ ID NO:48) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Env.

10 Figure 20 (SEQ ID NO:49) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 21 (SEQ ID NO:50) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Env gp160.

15 Figure 22 (SEQ ID NO:51) depicts the nucleotide sequence of a synthetic Gag-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 23 (SEQ ID NO:52) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Gag.

Figure 24 (SEQ ID NO:53) depicts the nucleotide sequence of a synthetic Gag-encoding polynucleotide (major homology region) derived from 8_5_TV1_C.ZA.

20 Figure 25 (SEQ ID NO:54) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Gag major homology region.

Figure 26 (SEQ ID NO:55) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 8_5_TV1_C.ZA.

25 Figure 27 (SEQ ID NO:56) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Nef.

Figure 28 (SEQ ID NO:57) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence includes a mutation at position 125 which results in a non-functional gene product.

30 Figure 29 (SEQ ID NO:58) depicts the nucleotide sequence of a synthetic RNaseH-encoding polynucleotide derived from 8_5_TV1_C.ZA. RnaseH is a functional domain of the Pol gene, corresponding to p15 (Table A).

Figure 30 (SEQ ID NO:59) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA RNaseH.

Figure 31 (SEQ ID NO:60) depicts the nucleotide sequence of a synthetic integrase (Int)-encoding polynucleotide derived from 8_5_TV1_C.ZA. Int is a functional domain of 5 the Pol gene, corresponding to p31 (Table A).

Figure 32 (SEQ ID NO:61) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Int.

Figure 33 (SEQ ID NO:62) depicts the nucleotide sequence of a synthetic Pol-encoding polynucleotide derived from 8_5_TV1_C.ZA.

10 Figure 34 (SEQ ID NO:63) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Pol.

Figure 35 (SEQ ID NO:64) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide derived from 8_5_TV1_C.ZA.

15 Figure 36 (SEQ ID NO:65) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Prot.

Figure 37 (SEQ ID NO:66) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide derived from 8_5_TV1_C.ZA containing a mutation in which results in inactivation of the protease.

20 Figure 38 (SEQ ID NO:67) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA inactivated Prot.

Figure 39 (SEQ ID NO:68) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide and a synthetic reverse transcriptase (RT)-encoding polynucleotide, both derived from 8_5_TV1_C.ZA. The Prot and RT sequences both contain a mutation which results in inactivation of the gene product.

25 Figure 40 (SEQ ID NO:69) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA inactivated Prot/mutated RT.

Figure 41 (SEQ ID NO:70) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide and a synthetic reverse transcriptase (RT)-encoding polynucleotide, both derived from 8_5_TV1_C.ZA.

30 Figure 42 (SEQ ID NO:71) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Prot and RT.

Figure 43 (SEQ ID NO:72) depicts the nucleotide sequence of a synthetic rev-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 1 of rev. Wild-type rev has two exons.

5 Figure 44 (SEQ ID NO:73) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 1 of Rev.

Figure 45 (SEQ ID NO:74) depicts the nucleotide sequence of a synthetic rev-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 2 of rev.

10 Figure 46 (SEQ ID NO:75) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 2 of Rev.

Figure 47 (SEQ ID NO:76) depicts the nucleotide sequence of a synthetic RT-encoding polynucleotide derived from 8_5_TV1_C.ZA.

15 Figure 48 (SEQ ID NO:77) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA RT.

Figure 49 (SEQ ID NO:78) depicts the nucleotide sequence of a synthetic RT-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic polynucleotide includes a mutation in the RT coding sequence which renders the gene product inactive.

Figure 50 (SEQ ID NO:79) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA RT including a mutation which inactivates the RT gene product.

20 Figure 51 (SEQ ID NO:80) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 1 of Tat and further includes a mutation that renders the Tat gene product non-functional. Wild-type Tat has two exons.

25 Figure 52 (SEQ ID NO:81) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 1 of Tat.

Figure 53 (SEQ ID NO:82) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 1 of Tat.

30 Figure 54 (SEQ ID NO:83) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 2 of Tat.

Figure 55 (SEQ ID NO:84) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 2 of Tat.

Figure 56 (SEQ ID NO:85) depicts the nucleotide sequence of a synthetic Vif-encoding polynucleotide derived from 8_5_TV1_C.ZA.

5 Figure 57 (SEQ ID NO:86) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Vif.

Figure 58 (SEQ ID NO:87) depicts the nucleotide sequence of a synthetic Vpr-encoding polynucleotide derived from 8_5_TV1_C.ZA.

10 Figure 59 (SEQ ID NO:88) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Vpr.

Figure 60 (SEQ ID NO:89) depicts the nucleotide sequence of a synthetic Vpu-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 61 (SEQ ID NO:90) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Vpu.

15 Figure 62 (SEQ ID NO:91) depicts the nucleotide sequence of a synthetic rev-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exons 1 and 2 of rev.

Figure 63 (SEQ ID NO:92) depicts the wild-type nucleotide sequence of exons 1 and 2 of rev derived from 8_5_TV1_C.ZA.

20 Figure 64 (SEQ ID NO:93) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic polynucleotide includes both exons 1 and 2 of Tat and further includes a mutation in exon 1 which renders the gene product non-functional.

25 Figure 65 (SEQ ID NO:94) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic polynucleotide includes both exons 1 and 2 of Tat.

Figure 66 (SEQ ID NO:95) depicts the wild-type nucleotide sequence of exons 1 and 2 of Tat derived from 8_5_TV1_C.ZA.

30 Figure 67 (SEQ ID NO:96) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence includes a mutation at position 125 which results in a non-functional gene product and a mutation that eliminates the myristylation site of the Nef gene product.

Figure 68 (SEQ ID NO:97) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 69 (SEQ ID NO:98) depicts the wild-type nucleotide sequence of Env gp160 derived from 12-5_1_TV2_C.ZA.

5 Figure 70 (SEQ ID NO:99) depicts the nucleotide sequence of a synthetic Gag-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 71 (SEQ ID NO:100) depicts the wild-type nucleotide sequence of Gag derived from 12-5_1_TV2_C.ZA.

10 Figure 72 (SEQ ID NO:101) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 73 (SEQ ID NO:102) depicts the wild-type nucleotide sequence of Nef derived from 12-5_1_TV2_C.ZA.

15 Figure 74 (SEQ ID NO:103) depicts the nucleotide sequence of a synthetic Pol-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 75 (SEQ ID NO:104) depicts the wild-type nucleotide sequence of Pol derived from 12-5_1_TV2_C.ZA.

20 Figure 76 (SEQ ID NO:105) depicts the nucleotide sequence of a synthetic Rev-encoding polynucleotide derived from exon 1 of Rev from 12-5_1_TV2_C.ZA.

Figure 77 (SEQ ID NO:106) depicts the wild-type nucleotide sequence of exon 1 of Rev derived from 12-5_1_TV2_C.ZA.

25 Figure 78 (SEQ ID NO:107) depicts the nucleotide sequence of a synthetic Rev-encoding polynucleotide derived from exon 2 of Rev from 12-5_1_TV2_C.ZA.

Figure 79 (SEQ ID NO:108) depicts the wild-type nucleotide sequence of exon 2 of Rev derived from 12-5_1_TV2_C.ZA.

25 Figure 80 (SEQ ID NO:109) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from exon 1 of Tat from 12-5_1_TV2_C.ZA.

Figure 81 (SEQ ID NO:110) depicts the wild-type nucleotide sequence of exon 1 of Tat derived from 12-5_1_TV2_C.ZA.

30 Figure 82 (SEQ ID NO:111) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from exon 2 of Tat from 12-5_1_TV2_C.ZA.

Figure 83 (SEQ ID NO:112) depicts the wild-type nucleotide sequence of exon 2 of Tat derived from 12-5_1_TV2_C.ZA.

Figure 84 (SEQ ID NO:113) depicts the nucleotide sequence of a synthetic Vif-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 85 (SEQ ID NO:114) depicts the wild-type nucleotide sequence of Vif derived from 12-5_1_TV2_C.ZA.

5 Figure 86 (SEQ ID NO:115) depicts the nucleotide sequence of a synthetic Vpr-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 87 (SEQ ID NO:116) depicts the wild-type nucleotide sequence of Vpr derived from 12-5_1_TV2_C.ZA.

10 Figure 88 (SEQ ID NO:117) depicts the nucleotide sequence of a synthetic Vpu-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 89 (SEQ ID NO:118) depicts the wild-type nucleotide sequence of Vpu derived from 12-5_1_TV2_C.ZA.

Figure 90 (SEQ ID NO:119) depicts the nucleotide sequence of a synthetic Env gp120-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted.

15 The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a codon modified signal peptide leader sequence (nucleotides 7 to 87); a gp120 coding sequence (nucleotides 88 to 1464); a stop codon (nucleotides 1465 to 1467); an XhoI restriction site (nucleotides 1468 to 1473).

20 Figure 91 (SEQ ID NO:120) depicts the nucleotide sequence of a synthetic Env gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp140 coding sequence (nucleotides 88 to 1977); a stop codon (nucleotides 1978 to 1980); an XhoI restriction site (nucleotides 1981 to 1986).

25 Figure 92 (SEQ ID NO:121) depicts the nucleotide sequence of a synthetic Env gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted and the sequence includes mutations in the cleavage site that prevent the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); gp140 coding sequence (nucleotides 88 to 1977); a stop codon (nucleotides 1978 to 1980); an XhoI restriction site (nucleotides 1981 to 1986).

Figure 93 (SEQ ID NO:122) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V1/V2 regions are deleted. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); gp160 coding sequence (nucleotides 88 to 2388); a stop codon (nucleotides 2389 to 2391); an XhoI restriction site (nucleotides 2392 to 2397).

Figure 94 (SEQ ID NO:123) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 88 to 2520); a stop codon (nucleotides 2521 to 2523); an XhoI restriction site (nucleotides 2524 to 2529).

Figure 95 (SEQ ID NO:124) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted and the cleavage site is mutated. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 88 to 2520); a stop codon (nucleotides 2521 to 2523); an XhoI restriction site (nucleotides 2524 to 2529).

Figure 96 (SEQ ID NO:125) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes a TPA1 leader sequence. The sequence includes: a SalI restriction site (nucleotides 1 to 6); a Kozak sequence (nucleotides 7 to 12); a TPA1 signal peptide leader sequence (nucleotides 13 to 87); a gp160 coding sequence (nucleotides 88 to 2604); a stop codon (nucleotides 2605 to 2607); an XhoI restriction site (nucleotides 2608 to 2613).

Figure 97 (SEQ ID NO:126) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 8 to 2607); a stop codon (nucleotides 2608 to 2610); an XhoI restriction site (nucleotides 2611 to 2616).

Figure 98 (SEQ ID NO:127) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes a wild type leader sequence. The sequence includes: an EcoRI restriction site

(nucleotides 1 to 6); a native (unmodified) signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 88 to 2607); a stop codon (nucleotides 2608 to 2610); an XhoI restriction site (nucleotides 2611 to 2616).

Figure 99 (SEQ ID NO:128) depicts the nucleotide sequence of wild type gp160 derived from 8_2_TV1_C.ZA.

Figure 100 (SEQ ID NO:131) depicts the nucleotide sequence of a synthetic Env gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes a TPA1 leader sequence (nucleotides 1-75); a gp140 coding sequence (nucleotides 76 to 2049); a stop codon (nucleotides 2050 to 2052)

Figure 101 (SEQ ID NO:132) depicts the nucleotide sequence of a synthetic gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes an EcoRI restriction site (nucleotides 1 to 6); a leader sequence modified from the TV1_C.ZA wild-type leader sequence (nucleotides 7 to 87); a gp140 coding sequence (nucleotides 88 to 2064); a stop codon (nucleotides 2065 to 2067); a XhoI restriction site (nucleotides 2068 to 2073).

Figure 102 (SEQ ID NO:133) depicts the nucleotide sequence of a synthetic gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes wild-type TV1_C.ZA unmodified leader sequence. The nucleotide sequence includes a restriction site (nucleotides 1 to 6); a wild type leader sequence (nucleotides 7 to 87); a gp140 coding sequence (nucleotides 88 to 2064); a stop codon (nucleotides 2065 to 2067); a XhoI restriction site (nucleotides 2068-2073).

Figure 103 (SEQ ID NO:134) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 12-5_1_TV2_C.ZA. The sequence includes a mutation at position 125 which results in a non-functional gene product.

Figure 104 (SEQ ID NO:135) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 12-5_1_TV2_C.ZA. The synthetic polynucleotide includes a mutation that eliminates the myristoylation site of the Nef gene product.

Figure 105 depicts an alignment of Env polypeptides from various HIV isolates. The regions between the arrows indicate regions (of TV1 and TV2 clones) in the beta and/or bridging sheet region(s) that can be deleted and/or truncated. The "*" denotes N-linked glycosylation sites (of TV1 and TV2 clones), one or more of which can be modified (e.g., deleted and/or mutated).

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Short Protocols in Molecular Biology*, 4th ed: (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

1. DEFINITIONS

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

"Synthetic" sequences, as used herein, refers to Type C HIV polypeptide-encoding polynucleotides whose expression has been modified as described herein, for example, by codon substitution and inactivation of inhibitory sequences. "Wild-type" or "native" sequences, as used herein, refers to polypeptide encoding sequences that are essentially as they are found in nature, e.g., Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env and/or Nef encoding sequences as found in Type C isolates, e.g., AF110965, AF110967, AF110968, AF110975, 8_5_TV1_C.ZA, 8_2_TV1_C.ZA or 12-5_1_TV2_C.ZA. The various regions of the HIV genome are shown in Table A, with numbering relative to 8_5_TV1_C.ZA (SEQ ID NO:33). Thus, the term "Pol" refers to one or more of the following polypeptides: polymerase (p6Pol); protease (prot); reverse transcriptase (p66RT or RT); RNaseH (p15RNaseH); and/or integrase (p31Int or Int).

As used herein, the term "virus-like particle" or "VLP" refers to a nonreplicating, viral shell, derived from any of several viruses discussed further below. VLPs are generally composed of one or more viral proteins, such as, but not limited to those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an appropriate expression system. Methods for producing particular VLPs are known in the art and discussed more fully below. The presence of VLPs following recombinant expression of viral proteins can be detected using conventional techniques known in the art, such as by electron microscopy, X-ray crystallography, and the like. See, e.g., Baker et al., *Biophys. J.* (1991) 60:1445-1456; Hagensee et al., *J. Virol.* 1994) 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation and/or identified by characteristic density banding. Alternatively, cryoelectron microscopy can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

By "particle-forming polypeptide" derived from a particular viral protein is meant a full-length or near full-length viral protein, as well as a fragment thereof, or a viral protein with internal deletions, which has the ability to form VLPs under conditions that favor VLP formation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, truncated and partial sequences, as well as analogs and precursor forms of the reference molecule. The term therefore intends deletions, additions and substitutions to the sequence, so long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also includes deletions, additions and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP. Preferred substitutions are those which are conservative in nature, i.e., those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cystine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids.

An "antigen" refers to a molecule containing one or more epitopes (either linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen." Normally, a B-cell epitope will include at least about 5 amino acids but can 5 be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a helper T-cell epitope at least about 12-20 amino acids.

Normally, an epitope will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (i.e., antigens which are separate and discrete from a whole organism with which the antigen is associated in nature), 10 as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes. Antibodies such as anti-idiotype antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or 15 polynucleotide which expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the definition of antigen herein.

For purposes of the present invention, antigens can be derived from any of several known viruses, bacteria, parasites and fungi, as described more fully below. The term also intends any of the various tumor antigens. Furthermore, for purposes of the present 20 invention, an "antigen" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the ability to elicit an immunological response, as defined herein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the antigens.

25 An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important 30 aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTL"s). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the

surfaces of cells. CTLs help induce and promote the destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376. Recent methods of measuring cell-mediated immune response include measurement of intracellular cytokines or cytokine secretion by T-cell populations, or by measurement of epitope specific T-cells (e.g., by the tetramer technique)(reviewed by McMichael, A.J., and O'Callaghan, C.A., *J. Exp. Med.* 187(9):1367-1371, 1998; Mcheyzer-Williams, M.G., et al, *Immunol. Rev.* 150:5-21, 1996; Lalvani, A., et al, *J. Exp. Med.* 186:859-865, 1997).

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or $\gamma\delta$ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent

cell cytotoxicity (ADCC) to provide protection to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

An "immunogenic composition" is a composition that comprises an antigenic molecule where administration of the composition to a subject results in the development in 5 the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest. The immunogenic composition can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal (e.g., intra-rectally or intra-vaginally) administration.

By "subunit vaccine" is meant a vaccine composition which includes one or more 10 selected antigens but not all antigens, derived from or homologous to, an antigen from a pathogen of interest such as from a virus, bacterium, parasite or fungus. Such a composition is substantially free of intact pathogen cells or pathogenic particles, or the lysate of such cells or particles. Thus, a "subunit vaccine" can be prepared from at least partially purified 15 (preferably substantially purified) immunogenic polypeptides from the pathogen, or analogs thereof. The method of obtaining an antigen included in the subunit vaccine can thus include standard purification techniques, recombinant production, or synthetic production.

"Substantially purified" general refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a 20 substantially purified component comprises 50%, preferably 80%-85%, more preferably 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a 25 nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from viral, prokaryotic 30 or eucaryotic mRNA, genomic DNA sequences from viral or prokaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence such as a stop codon may be located 3' to the coding sequence.

Typical "control elements", include, but are not limited to, transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences (located 3' to the translation stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), and translation termination sequences.

5 A "polynucleotide coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. Exemplary coding sequences are the modified viral polypeptide-coding sequences of the present invention. A transcription termination sequence may be located 3' to the coding sequence. Typical "control elements", include, but are not limited to, transcription regulators, such as promoters, transcription enhancer elements, transcription termination signals, and polyadenylation sequences; and translation regulators, such as sequences for optimization of initiation of translation, *e.g.*, Shine-Dalgarno (ribosome binding site) sequences, Kozak sequences (*i.e.*, sequences for the optimization of translation, located, for example, 5' to the coding sequence), leader sequences, translation initiation codon (*e.g.*, ATG), and translation termination sequences. In certain embodiments, one or more translation regulation or initiation sequences (*e.g.*, the leader sequence) are derived from wild-type translation initiation sequences, *i.e.*, sequences that regulate translation of the coding region in their native state. Wild-type leader sequences that have been modified, using the methods described herein, also find use in the present invention. Promoters can include inducible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), repressible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), and constitutive promoters.

10 15 20 25 30 A "nucleic acid" molecule can include, but is not limited to, prokaryotic sequences, eucaryotic mRNA, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (*e.g.*, mammalian) DNA, and even synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter

operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the 5 promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with 10 which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting prokaryotic microorganisms or eukaryotic cell lines cultured as unicellular 15 entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the 20 parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or 25 more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then can be determined between the compared polypeptide sequences.

Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of the mRNA for that gene 30 (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to

an exact nucleotide to nucleotide or amino acid to amino acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981).

This algorithm can be extended to use with peptide sequences using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure*, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program BLAST, which can also be used with default parameters. For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand =

both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address: <http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

5 One of skill in the art can readily determine the proper search parameters to use for a given sequence, exemplary preferred Smith Waterman based parameters are presented above. For example, the search parameters may vary based on the size of the sequence in question. Thus, for the polynucleotide sequences of the present invention the length of the polynucleotide sequence disclosed herein is searched against a selected database and 10 compared to sequences of essentially the same length to determine percent identity. For example, a representative embodiment of the present invention would include an isolated polynucleotide having X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least about a selected level of percent identity relative to Y contiguous nucleotides of the sequences described herein, and (ii) for search purposes X equals Y, wherein Y is a 15 selected reference polynucleotide of defined length.

The sequences of the present invention can include fragments of the sequences, for example, from about 15 nucleotides up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing, Figures, and claims), including all integer values falling within the above-described range. For example, fragments of the 20 polynucleotide sequences of the present invention may be 30-60 nucleotides, 60-120 nucleotides, 120-240 nucleotides, 240-480 nucleotides, 480-1000 nucleotides, and all integer values therebetween.

25 The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%, greater than 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% up to 100% (including all integer values falling within these described ranges) sequence identity to the synthetic expression cassette (and purified 30 polynucleotide) sequences disclosed herein (for example, to the claimed sequences or other sequences of the present invention) when the sequences of the present invention are used as the query sequence against, for example, a database of sequences.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the

efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g.,

5 Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *supra* or Ausubel et al., *supra*). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe

10 having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions.

f a particular set of hybridization conditions is selected following standard art (see, for example, Sambrook, et al., *supra* or Ausubel et al., *supra*).

Polynucleotide is "derived from" second polynucleotide if it has the same or e same basepair sequence as a region of the second polynucleotide, its cDNA, iereof, or if it displays sequence identity as described above.

Polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first derived from a second polynucleotide, or (ii) displays sequence identity to the tides as described above.

lly, a viral polypeptide is "derived from" a particular polypeptide of a virus ide) if it is (i) encoded by an open reading frame of a polynucleotide of that ynucleotide), or (ii) displays sequence identity to polypeptides of that virus as e.

"Encoded by" refers to a nucleic acid sequence which codes for a polypeptide rein the polypeptide sequence or a portion thereof contains an amino acid least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even ly at least 15 to 20 amino acids from a polypeptide encoded by the nucleic acid o encompassed are polypeptide sequences which are immunologically ith a polypeptide encoded by the sequence. Further, polyproteins can be , fusing in-frame two or more polynucleotide sequences encoding polypeptide ducts. Further, polycistronic coding sequences may be produced by placing olynuclotide sequences encoding polypeptide products adjacent each other, r the control of one promoter, wherein each polypeptide coding sequence may o include sequences for internal ribosome binding sites.

"Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof initially free, e.g., contains less than about 50%, preferably less than about 70%, ferably less than about 90%, of the protein with which the polynucleotide is ciated. Techniques for purifying polynucleotides of interest are well-known in clude, for example, disruption of the cell containing the polynucleotide with a ent and separation of the polynucleotide(s) and proteins by ion-exchange hy, affinity chromatography and sedimentation according to density.

"Nucleic acid immunization" is meant the introduction of a nucleic acid molecule or more selected antigens into a host cell, for the *in vivo* expression of an

antigen, antigens, an epitope, or epitopes. The nucleic acid molecule can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter case, the transformed cells are reintroduced into the subject where
5 an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

"Gene transfer" or "gene delivery" refers to methods or systems for reliably inserting DNA of interest into a host cell. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred
10 replicons (e.g., episomes), or integration of transferred genetic material into the genomic DNA of host cells. Gene delivery expression vectors include, but are not limited to, vectors derived from alphaviruses, pox viruses and vaccinia viruses. When used for immunization, such gene delivery expression vectors may be referred to as vaccines or vaccine vectors.

"T lymphocytes" or "T cells" are non-antibody producing lymphocytes that constitute
15 a part of the cell-mediated arm of the immune system. T cells arise from immature lymphocytes that migrate from the bone marrow to the thymus, where they undergo a maturation process under the direction of thymic hormones. Here, the mature lymphocytes rapidly divide increasing to very large numbers. The maturing T cells become
20 immunocompetent based on their ability to recognize and bind a specific antigen. Activation of immunocompetent T cells is triggered when an antigen binds to the lymphocyte's surface receptors.

The term "transfection" is used to refer to the uptake of foreign DNA by a cell. A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. See, e.g., Graham et al.
25 (1973) *Virology*, 52:456, Sambrook et al. (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, and Chu et al. (1981) *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells. The term refers to both stable and transient uptake of the genetic material, and includes uptake of peptide- or
30 antibody-linked DNAs.

A "vector" is capable of transferring gene sequences to target cells (e.g., viral vectors, non-viral vectors, particulate carriers, and liposomes). Typically, "vector construct,"

"expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Transfer of a "suicide gene" (e.g., a drug-susceptibility gene) to a target cell renders the cell sensitive to compounds or compositions that are relatively nontoxic to normal cells.

5 Moolten, F.L. (1994) *Cancer Gene Ther.* 1:279-287. Examples of suicide genes are thymidine kinase of herpes simplex virus (HSV-tk), cytochrome P450 (Manome et al. (1996) *Gene Therapy* 3:513-520), human deoxycytidine kinase (Manome et al. (1996) *Nature Medicine* 2(5):567-573) and the bacterial enzyme cytosine deaminase (Dong et al. (1996) *Human Gene Therapy* 7:713-720). Cells which express these genes are rendered sensitive to the effects of the relatively nontoxic prodrugs ganciclovir (HSV-tk), cyclophosphamide (cytochrome P450 2B1), cytosine arabinoside (human deoxycytidine kinase) or 5-fluorocytosine (bacterial cytosine deaminase). Culver et al. (1992) *Science* 256:1550-1552, Huber et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:8302-8306.

10

15 A "selectable marker" or "reporter marker" refers to a nucleotide sequence included in a gene transfer vector that has no therapeutic activity, but rather is included to allow for simpler preparation, manufacturing, characterization or testing of the gene transfer vector.

A "specific binding agent" refers to a member of a specific binding pair of molecules wherein one of the molecules specifically binds to the second molecule through chemical and/or physical means. One example of a specific binding agent is an antibody directed 20 against a selected antigen.

25 By "subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The system described above is intended for use in any of the above vertebrate species, since 30 the immune systems of all of these vertebrates operate similarly.

By "pharmaceutically acceptable" or "pharmacologically acceptable" is meant a material which is not biologically or otherwise undesirable, i.e., the material may be

administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the composition in which it is contained.

By "physiological pH" or a "pH in the physiological range" is meant a pH in the range 5 of approximately 7.2 to 8.0 inclusive, more typically in the range of approximately 7.2 to 7.6 inclusive.

As used herein, "treatment" refers to any of (I) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may be 10 effected prophylactically (prior to infection) or therapeutically (following infection).

By "co-administration" is meant administration of more than one composition or molecule. Thus, co-administration includes concurrent administration or sequentially administration (in any order), via the same or different routes of administration. Non-limiting examples of co-administration regimes include, co-administration of nucleic acid and 15 polypeptide; co-administration of different nucleic acids (e.g., different expression cassettes as described herein and/or different gene delivery vectors); and co-administration of different polypeptides (e.g., different HIV polypeptides and/or different adjuvants). The term also encompasses multiple administrations of one of the co-administered molecules or compositions (e.g., multiple administrations of one or more of the expression cassettes 20 described herein followed by one or more administrations of a polypeptide-containing composition). In cases where the molecules or compositions are delivered sequentially, the time between each administration can be readily determined by one of skill in the art in view of the teachings herein.

"Lentiviral vector", and "recombinant lentiviral vector" refer to a nucleic acid 25 construct which carries, and within certain embodiments, is capable of directing the expression of a nucleic acid molecule of interest. The lentiviral vector include at least one transcriptional promoter/enhancer or locus defining element(s), or other elements which control gene expression by other means such as alternate splicing, nuclear RNA export, post-translational modification of messenger, or post-transcriptional modification of protein. Such 30 vector constructs must also include a packaging signal, long terminal repeats (LTRS) or portion thereof, and positive and negative strand primer binding sites appropriate to the retrovirus used (if these are not already present in the retroviral vector). Optionally, the

recombinant lentiviral vector may also include a signal which directs polyadenylation, selectable markers such as Neo, TK, hygromycin, phleomycin, histidinol, or DHFR, as well as one or more restriction sites and a translation termination sequence. By way of example, such vectors typically include a 5'LTR, a tRNA binding site, a packaging signal, an origin of second strand DNA synthesis, and a 3'LTR or a portion thereof.

"Lentiviral vector particle" as utilized within the present invention refers to a lentivirus which carries at least one gene of interest. The retrovirus may also contain a selectable marker. The recombinant lentivirus is capable of reverse transcribing its genetic material (RNA) into DNA and incorporating this genetic material into a host cell's DNA upon infection. Lentiviral vector particles may have a lentiviral envelope, a non-lentiviral envelope (e.g., an amphi or VSV-G envelope), or a chimeric envelope.

"Nucleic acid expression vector" or "Expression cassette" refers to an assembly which is capable of directing the expression of a sequence or gene of interest. The nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well. Expression cassettes described herein may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include a bacterial origin of replication, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

"Packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant retrovirus which are lacking in a recombinant retroviral vector. Typically, such packaging cells contain one or more expression cassettes which are capable of expressing proteins which encode *Gag*, *pol* and *env* proteins.

"Producer cell" or "vector producing cell" refers to a cell which contains all elements necessary for production of recombinant retroviral vector particles.

2. MODES OF CARRYING OUT THE INVENTION

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

5 **2.1. THE HIV GENOME**

The HIV genome and various polypeptide-encoding regions are shown in Table A. The nucleotide positions are given relative to 8_5_TV1_C.ZA (SEQ ID NO:33, Figure 11). However, it will be readily apparent to one of ordinary skill in the art in view of the teachings of the present disclosure how to determine corresponding regions in other HIV strains or variants (e.g., isolates HIV_{IIIb}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}), and simian immunodeficiency virus (SIV). (See, e.g., *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA; for a description of these and other related viruses), using for example, sequence comparison programs (e.g., BLAST and others described herein) or identification and alignment of structural features (e.g., a program such as the "ALB" program described herein that can identify the various regions).

Table A: Regions of the HIV Genome relative to 8_5_TV1_C.ZA

<u>Region</u>	<u>Position in nucleotide sequence</u>
5'LTR	1-636
U3	1-457
5 R	458-553
U5	554-636
NFkB II	340-348
NFkB I	354-362
Sp1 III	379-388
10 Sp1 II	390-398
Sp1 I	400-410
TATA Box	429-433
TAR	474-499
Poly A signal	529-534
15 PBS	638-655
p7 binding region, packaging signal	685-791
20 Gag:	792-2285
p17	792-1178
p24	1179-1871
Cyclophilin A bdg.	1395-1505
MHR	1632-1694
25 p2	1872-1907
p7	1908-2072
Frameshift slip	2072-2078
p1	2073-2120
p6Gag	2121-2285
30 Zn-motif I	1950-1991
Zn-motif II	2013-2054

	Pol:	2072-5086
	p6Pol	2072-2245
	Prot	2246-2542
	p66RT	2543-4210
5	p15RNaseH	3857-4210
	p31Int	4211-5086
	 Vif:	 5034-5612
	Hydrophilic region	5292-5315
10	 Vpr:	 5552-5839
	Oligomerization	5552-5677
	Amphipathic α -helix	5597-5653
15	 Tat:	 5823-6038 and 8417-8509
	Tat-1 exon	5823-6038
	Tat-2 exon	8417-8509
	N-terminal domain	5823-5885
	Trans-activation domain	5886-5933
20	Transduction domain	5961-5993
	 Rev:	 5962-6037 and 8416-8663
	Rev-1 exon	5962-6037
	Rev-2 exon	8416-8663
25	High-affinity bdg. site	8439-8486
	Leu-rich effector domain	8562-8588
	 Vpu:	 6060-6326
	Transmembrane domain	6060-6161
30	Cytoplasmic domain	6162-6326

	Env (gp160):	6244-8853
	Signal peptide	6244-6324
	gp120	6325-7794
	V1	6628-6729
5	V2	6727-6852
	V3	7150-7254
	V4	7411-7506
	V5	7663-7674
	C1	6325-6627
10	C2	6853-7149
	C3	7255-7410
	C4	7507-7662
	C5	7675-7794
	CD4 binding	7540-7566
15	gp41	7795-8853
	Fusion peptide	7789-7842
	Oligomerization domain	7924-7959
	N-terminal heptad repeat	7921-8028
	C-terminal heptad repeat	8173-8280
20	Immunodominant region	8023-8076
	 Nef:	 8855-9478
	Myristoylation	8858-8875
	SH3 binding	9062-9091
25	Polypurine tract	9128-9154
	SH3 binding	9296-9307

It will be readily apparent that one of skill in the art can readily align any sequence to that shown in Table A to determine relative locations of any particular HIV gene. For example, using one of the alignment programs described herein (e.g., BLAST), other HIV Type C sequences can be aligned with 8_5_TV1_C.ZA (Table A) and locations of genes determined.

Polypeptide sequences can be similarly aligned. For example, Figure 103 shows the alignment of Env polypeptide sequences from various strains, relative to SF-162. As described in detail in co-owned WO/39303, Env polypeptides (e.g., gp120, gp140 and gp160) include a "bridging sheet" comprised of 4 anti-parallel β -strands (β -2, β -3, β -20 and β -21) that form a β -sheet. Extruding from one pair of the β -strands (β -2 and β -3) are two loops, V1

and V2. The β -2 sheet occurs at approximately amino acid residue 113 (Cys) to amino acid residue 117 (Thr) while β -3 occurs at approximately amino acid residue 192 (Ser) to amino acid residue 194 (Ile), relative to SF-162 (see, Figure 103). The "V1/V2 region" occurs at approximately amino acid positions 120 (Cys) to residue 189 (Cys), relative to SF-162.

- 5 Extruding from the second pair of β -strands (β -20 and β -21) is a "small-loop" structure, also referred to herein as "the bridging sheet small loop." The locations of both the small loop and bridging sheet small loop can be determined relative to HXB-2 following the teachings herein and in WO/39303. Also shown by arrows in Figure 103A-C are approximate sites for deletions sequence from the beta sheet region. The "*" denotes N-glycosylation sites that can
10 be mutated following the teachings of the present specification.

2.2 SYNTHETIC EXPRESSION CASSETTES

2.2.1 MODIFICATION OF HIV-1-TYPE C *POL*-, *PROT*-, *RT*-, *INT*-, *GAG*, *ENV*, *TAT*, *REV*, *NEF*, *RNASEH*, *VIF*, *VPR*, AND *VPU* NUCLEIC ACID CODING SEQUENCES

- 15 One aspect of the present invention is the generation of HIV-1 type C coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequences.

2.2.1.1 MODIFICATION OF *GAG* NUCLEIC ACID CODING SEQUENCES

- 20 An exemplary embodiment of the present invention is illustrated herein by modifying the Gag protein wild-type sequences obtained from the AF110965 and AF110967 strains of HIV-1, subtype C. (see, for example, Korber et al. (1998)*Human Retroviruses and Aids*, Los Alamos, New Mexico: Los Alamos National Laboratory; Novitsky et al. (1999) *J. Virol.* 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). Also illustrated herein is the modification of wild-type sequences from novel isolates 8_5_TV1_C.ZA (also called TV001 or TV1) and 12-5_1_TV2_C.ZA (also called TV002 or TV2). SEQ ID NO:52 shows the wild-type sequence of Gag from 8_5_TV1_C.ZA and SEQ ID NO:54 shows the wild-type sequence of the major homology region of Gag (nucleotides 1632-1694 of Table A) of the same strain. SEQ ID NO:100
25 shows the wild-type sequence of Gag of 12-5_1_TV2_C.ZA.
30

Gag sequence obtained from other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Gag protein encoding sequences obtained from the isolates of HIV-1 Type C, for example as described in Novitsky et al., (1999), *supra*; Myers et al., *infra*; 5 Virology, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA and on the World Wide Web (Internet), for example at <http://hiv-web.lan1.gov/cgi-bin/hivDB3/public/wdb/ssampublic> and <http://hiv-web.lan1.gov>.

10 First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In 15 comparison, highly expressed human codons prefer the nucleotides G or C. The Gag coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag coding sequences. The RRE is a secondary RNA structure that 20 interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements can be inactivated by introducing multiple point mutations that do not alter the reading frame of the encoded proteins.

Subtype C Gag-encoding sequences having inactivated RRE sites are shown, for example, in 25 Figures 1 (SEQ ID NO:3), 2 (SEQ ID NO:4), 5 (SEQ ID NO:20) and 6 (SEQ ID NO:26). Similarly, other synthetic polynucleotides derived from other Subtype C strains can be modified to inactivate the RRE sites.

Modification of the Gag polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as 30 other types of cell lines, including, but not limited to, insect cells). Further, expression of the sequences results in production of virus-like particles (VLPs) by these cell lines (see below).

2.2.1.2 MODIFICATION OF ENV NUCLEIC ACID CODING SEQUENCES

Similarly, the present invention also includes synthetic Env-encoding polynucleotides and modified Env proteins. Wild-type Env sequences are obtained from the AF110968 and AF110975 strains as well as novel strains 8_5_TV1_C.ZA (SEQ ID NO:33) and 12-

5 5_1_TV2_C.ZA (SEQ ID NO:45) of HIV-1, type C. (see, for example, Novitsky et al. (1999) *J. Virol.* 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). Wild-type Env sequences of 8_5_TV1_C.ZA are shown, for example, in SEQ ID NO:48 (wild-type Env common region, nucleotides 7486-7629 as shown in Table A); and SEQ ID NO:50 (wild type gp160, nucleotides 6244-8853 as shown in Table A). Wild-type
10 Env gp160 of 12-5_1_TV2_C.ZA is shown in SEQ ID NO:98. It will be readily apparent from the disclosure herein that polynucleotides encoding fragments of Env gp160 (e.g., gp120, gp41, gp140) can be readily obtained from the larger, full-length sequences disclosed herein. It will also be readily apparent that other modifications can be made, for example deletion of regions such as the V1 and/or V2 region; mutation of the cleavage site and the like
15 (see, Example 1). Exemplary sequences of such modification as shown in SEQ ID NO:119 through 127.

Further, Env sequences obtained from other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Env protein encoding sequences obtained from
20 the isolates of HIV-1 Type C, described above.

The codon usage pattern for Env was modified as described above for Gag so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. Experiments performed in support of the present invention show that the synthetic Env sequences were capable of higher level of protein production relative to
25 the native Env sequences.

Modification of the Env polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Similar Env polypeptide coding sequences can be obtained, modified and tested for improved expression from a
30 variety of isolates, including those described above for Gag.

Further modifications of Env include, but are not limited to, generating polynucleotides that encode Env polypeptides having mutations and/or deletions therein. For

instance, the hypervariable regions, V1 and/or V2, can be deleted as described herein. Additionally, other modifications, for example to the bridging sheet region and/or to N-glycosylation sites within Env can also be performed following the teachings of the present specification. (see, Figure 103A-C and WO/39303). Various combinations of these 5 modifications can be employed to generate synthetic expression cassettes as described herein.

2.2.1.3 MODIFICATION OF SEQUENCES INCLUDING HIV-1 *Pol* NUCLEIC ACID CODING SEQUENCES

The present invention also includes expression cassettes which include synthetic Pol 10 sequences. As noted above, "Pol" includes, but is not limited to, the protein-encoding regions shown in Figure 7, for example polymerase, protease, reverse transcriptase and/or integrase-containing sequences. The regions shown in Figure 7 are described, for example, in Wan et al (1996) *Biochem. J.* 316:569-573; Kohl et al. (1988) *PNAS USA* 85:4686-4690; Krausslich et al. (1988) *J. Virol.* 62:4393-4397; Coffin, "Retroviridae and their Replication" 15 in *Virology*, pp1437-1500 (Raven, New York, 1990); Patel et. al. (1995) *Biochemistry* 34:5351-5363. Thus, the synthetic expression cassettes exemplified herein include one or more of these regions and one or more changes to the resulting amino acid sequences.

Wild type Pol sequences were obtained from the AF110975, 8_5_TV1_C.ZA and 12-5_1_TV2_C.ZA strains of HIV-1, type C. (see, for example, Novitsky et al. (1999) *J. Virol.* 20 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). SEQ ID NO:34 shows the wild type sequence of AF110975 from the p2 through p7 region of Pol (see, Figure 7 and Table A). SEQ ID NO:35 shows the wild type sequence of AF110975 from p1 through the first 6 amino acids of integrase (see, Figure 7 and Table A). SEQ ID NO:63 and SEQ ID NO:104 show wild-type sequences of Pol from 8_5_TV1_C.ZA and 12-5_1_TV2_C.ZA, respectively (see, also, Table A).

Sequence obtained from other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Pol protein encoding sequences obtained from the isolates of HIV-1 Type C described herein.

30 The codon usage pattern for Pol was modified as described above for Gag and Env so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes.

Table B shows the nucleotide positions of various regions found in the Pol constructs exemplified herein (e.g., SEQ ID NOs: 30-32).

Table B

Region	Position in nucleotide sequence in construct		
	PR975(+) Seq Id No:30	PR975YM Seq Id No:31	PR975(+) YMWM Seq Id No:32
Sal 1 restriction site	1-6	1-6	1-6
Kozak start codon	7-16	7-16	7-16
p2	16-54	16-54	16-54
p7	55-219	55-219	55-219
p1/p6 pol	220-375	220-375	220-375
Insertion mutation for in frame	225	225	225
p10Protease	376-672	376-672	376-672
p66RT	673-2352	673-2346	673-2340
p51RT	673-1992	673-1986	673-1980
p15RNaseH	1993-2352	1993-2346	1993-2340
catalytic center region (YMDD)	1219-1230	1219-1224	1219-1224
primer grip region (WMGY)	1357-1368	1351-1362	1351-1356
6aa Integrase	2353-2370	2347-2364	2341-2358
YMDD epitope cassette (incl. 5'+3'Gly)	2371-2424	2365-2418	2359-2412
MCS (multiple cloning site)	2425-2463	2419-2457	2413-2451
EcoR 1 restriction site	2464-2469	2458-2463	2452-2457

As shown in Table B, exemplary constructs were modified in various ways. For example, the expression constructs exemplified herein include sequence that encodes the first 6 amino acids of the integrase polypeptide. This 6 amino acid region is believed to provide a cleavage recognition site recognized by HIV protease (see, e.g., McCornack et al. (1997) *FEBS Letts* 414:84-88). As noted above, certain constructs exemplified herein include a multiple cloning site (MCS) for insertion of one or more transgenes, typically at the 3' end of the construct. In addition, a cassette encoding a catalytic center epitope derived from the catalytic center in RT is typically included 3' of the sequence encoding 6 amino acids of integrase. This cassette (SEQ ID NO:36) encodes Ile178 through Serine 191 of RT (amino acids 3 through 16 of SEQ ID NO:37) and was added to keep this well conserved region as a possible CTL epitope. Further, the constructs contain an insertion mutations (position 225 of SEQ ID NOs:30 to 32) to preserve the reading frame. (see, e.g., Park et al. (1991) *J. Virol.* 65:5111).

In certain embodiments, the catalytic center and/or primer grip region of RT are modified. The catalytic center and primer grip regions of RT are described, for example, in Patel et al. (1995) *Biochem.* 34:5351 and Palaniappan et al. (1997) *J. Biol. Chem.* 272(17):11157. For example, in the construct designated PR975YM (SEQ ID NO:31), wild type sequence encoding the amino acids YMDD at positions 183-185 of p66 RT, numbered relative to AF110975, are replaced with sequence encoding the amino acids "AP". In the construct designated PR975YMWM (SEQ ID NO:32), the same mutation in YMDD is made and, in addition, the primer grip region (amino acids WMGY, residues 229-232 of p66RT, numbered relative to AF110975) are replaced with sequence encoding the amino acids "PI."

For the Pol sequence, the changes in codon usage are typically restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame; however, regions within the frameshift translation region can be modified as well. Finally, inhibitory (or instability) elements (INS) located within the coding sequences of the protease polypeptide coding sequence can be altered as well.

Experiments can be performed in support of the present invention to show that the synthetic Pol sequences were capable of higher level of protein production relative to the native Pol sequences. Modification of the Pol polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Similar Pol polypeptide coding sequences can be obtained, modified and tested for improved expression from a variety of isolates, including those described above for Gag and Env.

2.2.1.4 MODIFICATION OF OTHER HIV SEQUENCES

The present invention also includes expression cassettes which include synthetic HIV Type C sequences derived HIV genes other than Gag, Env and Pol, including but not limited to, regions within Gag, Env, Pol, as well as, vif, vpr, tat, rev, vpu, and nef, for example from 8_5_TV1_C.ZA (SEQ ID NO:33) or 12-5_1_TV2_C.ZA (SEQ ID NO:45). Sequences obtained from other strains can be manipulated in similar fashion following the teachings of the present specification.

As noted above, the codon usage pattern is modified as described above for Gag, Env and Pol so that the resulting nucleic acid coding sequence is comparable to codon usage found in highly expressed human genes. Experiments can be performed in support of the present invention to show that these synthetic sequences were capable of higher level of protein production relative to the native sequences and that modification of the wild-type polypeptide coding sequences results in improved expression relative to the wild-type coding

sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Furthermore, the nucleic acid sequence can also be modified to introduce mutations into one or more regions of the gene, for instance to render the gene product non-functional and/or to eliminate the myristylation site in Nef.

5 Synthetic expression cassettes exemplified herein include SEQ ID NO:49 and SEQ ID NO:97 (Env gp160-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:51 and SEQ ID NO:99 (Gag-encoding sequences modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:53 (Gag major homology region, modified based on 10 8_5_TV1_C.ZA wild type); SEQ ID NO:55 and SEQ ID NO:101 (Nef-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:57 and SEQ ID NO:134 (Nef-encoding sequences with a mutation at position 125 resulting in a non-functional gene product, modified based on 15 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA, respectively); SEQ ID NO:58 (RNaseH-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:60 (Integrase-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:62 and SEQ ID NO:103 (Pol-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:64 (Protease-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:66 (inactivated protease-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:68 (inactivated protease and RT mutated sequences, modified based on 20 8_5_TV1_C.ZA wild type); SEQ ID NO:70 (protease and reverse-transcriptase-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:72 and SEQ ID NO:105 (exon 1 of Rev, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:74 and SEQ ID NO:107 (exon 2 of Rev, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:76 (reverse transcriptase-encoding sequences, modified based on 25 8_5_TV1_C.ZA wild type); SEQ ID NO:78 (mutated reverse-transcriptase, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:80 (exon 1 of Tat including a mutation that results in non-functional Tat, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:81 and SEQ ID NO:109 (exon 1 of Tat, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:83 and SEQ ID NO:111 (exon 2 of Tat, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:85 and SEQ ID NO:113) (Vif-encoding sequences, modified 30 based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ 35

10 ID NO:87 and SEQ ID NO:115 (Vpr-encoding sequences, modified based on
8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:89
and SEQ ID NO:117 (Vpu-encoding sequences, modified based on 8_5_TV1_C.ZA wild
type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:91 (sequences of exons 1
5 and 2 of Rev, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:93 (sequences of
mutated exon 1 of Tat and exon 2 of Tat, where mutation of exon 1 results in non-functional
Tat, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:94 (sequences of exons 1
and 2 of Tat, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:96 and SEQ ID
NO:135 (Nef-encoding sequences including a mutation to eliminate myristylation site,
10 modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA, respectively).

2.2.1.5 FURTHER MODIFICATION OF SEQUENCES INCLUDING HIV-1 NUCLEIC ACID CODING SEQUENCES

The Type C HIV polypeptide-encoding expression cassettes described herein may
15 also contain one or more further sequences encoding, for example, one or more transgenes.
Further sequences (e.g., transgenes) useful in the practice of the present invention include, but
are not limited to, further sequences are those encoding further viral epitopes/antigens
{including but not limited to, HCV antigens (e.g., E1, E2; Houghton, M., et al., U.S. Patent
No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088,
20 issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November
4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J.,
et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No.
5,670,152, issued September 23, 1997), HIV antigens (e.g., derived from *tat*, *rev*, *nef* and/or
env); and sequences encoding tumor antigens/epitopes. Further sequences may also be
25 derived from non-viral sources, for instance, sequences encoding cytokines such interleukin-2
(IL-2), stem cell factor (SCF), interleukin 3 (IL-3), interleukin 6 (IL-6), interleukin 12 (IL-
12), G-CSF, granulocyte macrophage-colony stimulating factor (GM-CSF), interleukin-1
alpha (IL-1 α), interleukin-11 (IL-11), MIP-1 α , tumor necrosis factor (TNF), leukemia
inhibitory factor (LIF), c-kit ligand, thrombopoietin (TPO) and flt3 ligand, commercially
30 available from several vendors such as, for example, Genzyme (Framingham, MA),
Genentech (South San Francisco, CA), Amgen (Thousand Oaks, CA), R&D Systems and
Immunex (Seattle, WA). Additional sequences are described below, for example in Section
2.3. Also, variations on the orientation of the Gag and other coding sequences, relative to
each other, are described below.

HIV polypeptide coding sequences can be obtained from other Type C HIV isolates, see, e.g., Myers et al. Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory. Synthetic expression cassettes can be generated using such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1).

Further, the synthetic expression cassettes of the present invention include related polypeptide sequences having greater than 85%, preferably greater than 90%, more preferably greater than 95%, and most preferably greater than 98% sequence identity to the synthetic expression cassette sequences disclosed herein (for example, (SEQ ID NOS:30-32; SEQ ID NOS: 3, 4, 20, and 21 and SEQ ID NOS:5-17). Various coding regions are indicated in Figures 3 and 4, for example in Figure 3 (AF110968), nucleotides 1-81 (SEQ ID NO:18); nucleotides 82-1512 (SEQ ID NO:6) encode a gp120 polypeptide; nucleotides 1513 to 2547 (SEQ ID NO:10) encode a gp41 polypeptide; nucleotides 82-2025 (SEQ ID NO:7) encode a gp140 polypeptide and nucleotides 82-2547 (SEQ ID NO:8) encode a gp160 polypeptide. Similarly, in Figure 98 (SEQ ID NO:127, strain 8_2_TV1_C.ZA), nucleotides 1-6 are an EcoR1 restriction site; nucleotides 7-87 encode a wild-type (from 8_2_TV1_C.ZA) leader signal peptide; nucleotides 88 to 1563 encode a gp120 polypeptide; nucleotides 88 to 2064 encode a gp140 polypeptide; nucleotides 88 to 2607 encode a gp160 polypeptide.

2.2.3 EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1 SUBTYPE C AND RELATED POLYPEPTIDES

Synthetic HIV-encoding sequences (expression cassettes) of the present invention can be cloned into a number of different expression vectors to evaluate levels of expression and, in the case of Gag, production of VLPs. The synthetic DNA fragments for HIV polypeptides can be cloned into eucaryotic expression vectors, including, a transient expression vector, CMV-promoter-based mammalian vectors, and a shuttle vector for use in baculovirus expression systems. Corresponding wild-type sequences can also be cloned into the same vectors.

These vectors can then be transfected into a several different cell types, including a variety of mammalian cell lines (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines are then cultured under appropriate conditions and the levels of any appropriate polypeptide product can be evaluated in supernatants. (see, Table A and Example 2). For example, p24 can be used to evaluate Gag expression; gp160, gp140 or gp120 can be used to evaluate Env expression; p6pol can be used to evaluate Pol

expression; prot can be used to evaluate protease; p15 for RNaseH; p31 for Integrase; and other appropriate polypeptides for Vif, Vpr, Tat, Rev, Vpu and Nef. Further, modified polypeptides can also be used, for example, other Env polypeptides include, but are not limited to, for example, native gp160, oligomeric gp140, monomeric gp120 as well as modified and/or synthetic sequences of these polypeptides. The results of these assays demonstrate that expression of synthetic HIV polypeptide-encoding sequences are significantly higher than corresponding wild-type sequences.

Further, Western Blot analysis can be used to show that cells containing the synthetic expression cassette produce the expected protein at higher per-cell concentrations than cells containing the native expression cassette. The HIV proteins can be seen in both cell lysates and supernatants. The levels of production are significantly higher in cell supernatants for cells transfected with the synthetic expression cassettes of the present invention.

Fractionation of the supernatants from mammalian cells transfected with the synthetic expression cassette can be used to show that the cassettes provide superior production of HIV proteins and, in the case of Gag, VLPs, relative to the wild-type sequences.

Efficient expression of these HIV-containing polypeptides in mammalian cell lines provides the following benefits: the polypeptides are free of baculovirus contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Subtype C HIV-containing polypeptides in CHO cells which is not feasible in the absence of the increased expression obtained using the constructs of the present invention. Exemplary Mammalian cell lines include, but are not limited to, BHK, VERO, HT1080, 293, 293T, RD, COS-7, CHO, Jurkat, HUT, SUPT, C8166, MOLT4/clone8, MT-2, MT-4, H9, PM1, CEM, and CEMX174, such cell lines are available, for example, from the A.T.C.C.).

A synthetic Gag expression cassette of the present invention will also exhibit high levels of expression and VLP production when transfected into insect cells. Synthetic expression cassettes described herein also demonstrate high levels of expression in insect cells. Further, in addition to a higher total protein yield, the final product from the synthetic polypeptides consistently contains lower amounts of contaminating baculovirus proteins than the final product from the native Type C sequences.

Further, synthetic expression cassettes of the present invention can also be introduced into yeast vectors which, in turn, can be transformed into and efficiently expressed by yeast cells (*Saccharomyces cerevisea*; using vectors as described in Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998).

In addition to the mammalian and insect vectors, the synthetic expression cassettes of the present invention can be incorporated into a variety of expression vectors using selected expression control elements. Appropriate vectors and control elements for any given cell type can be selected by one having ordinary skill in the art in view of the teachings of the present specification and information known in the art about expression vectors.

For example, a synthetic expression cassette can be inserted into a vector which includes control elements operably linked to the desired coding sequence, which allow for the expression of the gene in a selected cell-type. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter (a CMV promoter can include intron A), RSV, HIV-Ltr, the mouse mammary tumor virus LTR promoter (MMLV-ltr), the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook, et al., *supra*, as well as a bovine growth hormone terminator sequence. Introns, containing splice donor and acceptor sites, may also be designed into the constructs for use with the present invention (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

The desired synthetic polypeptide encoding sequences can be cloned into any number of commercially available vectors to generate expression of the polypeptide in an appropriate host system. These systems include, but are not limited to, the following: baculovirus expression {Reilly, P.R., et al., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992); Beames, et al., *Biotechniques* 11:378 (1991); Pharmingen; Clontech, Palo Alto, CA)}, vaccinia expression {Earl, P. L., et al., "Expression of proteins in mammalian cells using vaccinia" In *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. Eds.),

Greene Publishing Associates & Wiley Interscience, New York (1991); Moss, B., *et al.*, U.S. Patent Number 5,135,855, issued 4 August 1992}, expression in bacteria {Ausubel, F.M., *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA; Clontech}, expression in yeast {Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998; Shuster, J.R., U.S. Patent No. 5,629,203, issued May 13, 1997; Gellissen, G., *et al.*, *Antonie Van Leeuwenhoek*, 62(1-2):79-93 (1992); Romanos, M.A., *et al.*, *Yeast* 8(6):423-488 (1992); Goeddel, D.V., *Methods in Enzymology* 185 (1990); Guthrie, C., and G.R. Fink, *Methods in Enzymology* 194 (1991)}, expression in mammalian cells {Clontech; Gibco-BRL, Ground Island, NY; e.g., Chinese hamster ovary (CHO) cell lines (Haynes, J., *et al.*, *Nuc. Acid. Res.* 11:687-706 (1983); 1983, Lau, Y.F., *et al.*, *Mol. Cell. Biol.* 4:1469-1475 (1984); Kaufman, R. J., "Selection and coamplification of heterologous genes in mammalian cells," in *Methods in Enzymology*, vol. 185, pp537-566. Academic Press, Inc., San Diego CA (1991)}, and expression in plant cells {plant cloning vectors, Clontech Laboratories, Inc., Palo Alto, CA, and Pharmacia LKB Biotechnology, Inc., Pistcataway, NJ; Hood, E., *et al.*, *J. Bacteriol.* 168:1291-1301 (1986); Nagel, R., *et al.*, *FEMS Microbiol. Lett.* 67:325 (1990); An, *et al.*, "Binary Vectors", and others in Plant Molecular Biology Manual A3:1-19 (1988); Miki, B.L.A., *et al.*, pp.249-265, and others in Plant DNA Infectious Agents (Hohn, T., *et al.*, eds.) Springer-Verlag, Wien, Austria, (1987); Plant Molecular Biology: Essential Techniques, P.G. Jones and J.M. Sutton, New York, J. Wiley, 1997; Miglani, Gurbachan *Dictionary of Plant Genetics and Molecular Biology*, New York, Food Products Press, 1998; Henry, R. J., *Practical Applications of Plant Molecular Biology*, New York, Chapman & Hall, 1997}.

Also included in the invention is an expression vector, containing coding sequences and expression control elements which allow expression of the coding regions in a suitable host. The control elements generally include a promoter, translation initiation codon, and translation and transcription termination sequences, and an insertion site for introducing the insert into the vector. Translational control elements have been reviewed by M. Kozak (e.g., Kozak, M., *Mamm. Genome* 7(8):563-574, 1996; Kozak, M., *Biochimie* 76(9):815-821, 1994; Kozak, M., *J Cell Biol* 108(2):229-241, 1989; Kozak, M., and Shatkin, A.J., *Methods Enzymol* 60:360-375, 1979).

Expression in yeast systems has the advantage of commercial production. Recombinant protein production by vaccinia and CHO cell line have the advantage of being mammalian expression systems. Further, vaccinia virus expression has several advantages including the following: (i) its wide host range; (ii) faithful post-transcriptional modification, processing, folding, transport, secretion, and assembly of recombinant proteins; (iii) high

level expression of relatively soluble recombinant proteins; and (iv) a large capacity to accommodate foreign DNA.

The recombinantly expressed polypeptides from synthetic HIV polypeptide-encoding expression cassettes are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including salt fractionation, ion exchange chromatography, gel filtration, size-exclusion chromatography, size-fractionation, and affinity chromatography. Immunoaffinity chromatography can be employed using antibodies generated based on, for example, HIV antigens.

Advantages of expressing the proteins of the present invention using mammalian cells include, but are not limited to, the following: well-established protocols for scale-up production; the ability to produce VLPs; cell lines are suitable to meet good manufacturing process (GMP) standards; culture conditions for mammalian cells are known in the art.

Various forms of the different embodiments of the invention, described herein, may be combined.

2.3 PRODUCTION OF VIRUS-LIKE PARTICLES AND USE OF THE CONSTRUCTS OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES.

The group-specific antigens (Gag) of human immunodeficiency virus type-1 (HIV-1) self-assemble into noninfectious virus-like particles (VLP) that are released from various eucaryotic cells by budding (reviewed by Freed, E.O., *Virology* 251:1-15, 1998). The synthetic expression cassettes of the present invention provide efficient means for the production of HIV-Gag virus-like particles (VLPs) using a variety of different cell types, including, but not limited to, mammalian cells.

Viral particles can be used as a matrix for the proper presentation of an antigen entrapped or associated therewith to the immune system of the host.

2.3.1 VLP PRODUCTION USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION

Experiments can be performed in support of the present invention to demonstrate that the synthetic expression cassettes of the present invention provide superior production of both Gag proteins and VLPs, relative to native Gag coding sequences. Further, electron microscopic evaluation of VLP production can show that free and budding immature virus particles of the expected size are produced by cells containing the synthetic expression cassettes.

Using the synthetic expression cassettes of the present invention, rather than native Gag coding sequences, for the production of virus-like particles provide several advantages. First, VLPs can be produced in enhanced quantity making isolation and purification of the VLPs easier. Second, VLPs can be produced in a variety of cell types using the synthetic expression cassettes, in particular, mammalian cell lines can be used for VLP production, for example, CHO cells. Production using CHO cells provides (i) VLP formation; (ii) correct myristylation and budding; (iii) absence of non-mammalian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification. The synthetic expression cassettes of the present invention are also useful for enhanced expression in cell-types other than mammalian cell lines. For example, infection of insect cells with baculovirus vectors encoding the synthetic expression cassettes results in higher levels of total Gag protein yield and higher levels of VLP production (relative to wild-type coding sequences). Further, the final product from insect cells infected with the baculovirus-Gag synthetic expression cassettes consistently contains lower amounts of contaminating insect proteins than the final product when wild-type coding sequences are used.

VLPs can spontaneously form when the particle-forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs produced using the synthetic expression cassettes of the present invention are conveniently prepared using recombinant techniques. As discussed below, the Gag polypeptide encoding synthetic expression cassettes of the present invention can include other polypeptide coding sequences of interest (for example, HIV protease, HIV polymerase, HCV core; Env; synthetic Env; see, Example 1). Expression of such synthetic expression cassettes yields VLPs comprising the Gag polypeptide, as well as, the polypeptide of interest.

Once coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Sambrook et al, *supra*. The vector is then used to transform an appropriate host cell. Suitable recombinant expression systems include, but are not limited to, bacterial, mammalian, baculovirus/insect, vaccinia, Semliki Forest virus (SFV), Alphaviruses (such as, Sindbis, Venezuelan Equine Encephalitis (VEE)), mammalian, yeast and Xenopus expression systems, well known in the art. Particularly preferred expression systems are mammalian cell lines, vaccinia, Sindbis, insect and yeast systems.

For example, a number of mammalian cell lines are known in the art and include immortalized cell lines available from the American Type Culture Collection (A.T.C.C.), such as, but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillermondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. See, e.g., Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987).

Viral vectors can be used for the production of particles in eucaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additionally, a vaccinia based infection/transfection system, as described in Tomei et al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. Alternately, T7 can be added as a purified protein or enzyme as in the "Progenitor" system (Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130). The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

Depending on the expression system and host selected, the VLPS are produced by growing host cells transformed by an expression vector under conditions whereby the particle-forming polypeptide is expressed and VLPs can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPs are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPs substantially intact. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated (or substantially purified) using methods that preserve the integrity thereof, such as, by gradient centrifugation, e.g., cesium chloride (CsCl) sucrose gradients, pelleting and the like (see, e.g., Kirnbauer et al. *J. Virol.* (1993) 67:6929-6936), as well as standard purification techniques including, e.g., ion exchange and gel filtration chromatography.

VLPs produced by cells containing the synthetic expression cassettes of the present invention can be used to elicit an immune response when administered to a subject. One advantage of the present invention is that VLPs can be produced by mammalian cells carrying the synthetic expression cassettes at levels previously not possible. As discussed above, the VLPs can comprise a variety of antigens in addition to the Gag polypeptide (e.g., Gag-protease, Gag-polymerase, Env, synthetic Env, etc.). Purified VLPs, produced using the synthetic expression cassettes of the present invention, can be administered to a vertebrate subject, usually in the form of vaccine compositions. Combination vaccines may also be used, where such vaccines contain, for example, an adjuvant subunit protein (e.g., Env). Administration can take place using the VLPs formulated alone or formulated with other antigens. Further, the VLPs can be administered prior to, concurrent with, or subsequent to, delivery of the synthetic expression cassettes for DNA immunization (see below) and/or delivery of other vaccines. Also, the site of VLP administration may be the same or different as other vaccine compositions that are being administered. Gene delivery can be accomplished by a number of methods including, but are not limited to, immunization with DNA, alphavirus vectors, pox virus vectors, and vaccinia virus vectors.

VLP immune-stimulating (or vaccine) compositions can include various excipients, adjuvants, carriers, auxiliary substances, modulating agents, and the like. The immune stimulating compositions will include an amount of the VLP/antigen sufficient to mount an immunological response. An appropriate effective amount can be determined by one of skill in the art. Such an amount will fall in a relatively broad range that can be determined through routine trials and will generally be an amount on the order of about 0.1 µg to about 1000 µg, more preferably about 1 µg to about 300 µg, of VLP/antigen.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-

glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen 5 may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc., as well as toxins derived from *E. coli*.

Adjuvants may also be used to enhance the effectiveness of the compositions. Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion 10 formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y 15 microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) RibiTM adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components 20 from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (3) saponin adjuvants, such as StimulonTM (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete 25 Freunds Adjuvant (CFA) and Incomplete Freunds Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (6) oligonucleotides or polymeric molecules encoding immunostimulatory CpG motifs (Davis, H.L., et al., *J. Immunology* 160:870-876, 1998; Sato, Y. et al., *Science* 273:352-354, 1996) or complexes of antigens/oligonucleotides {Polymeric molecules include double and single stranded RNA and DNA, and backbone modifications thereof, for example, 30 methylphosphonate linkages; or (7) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63) LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 35 (where lysine is substituted for the wild-type amino acid at position 9 and glycine substituted

at position 129) (see, e.g., International Publication Nos. W093/13202 and W092/19265); and (8) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Further, such polymeric molecules include alternative polymer backbone structures such as, but not limited to, polyvinyl backbones (Pitha, *Biochem Biophys Acta*, 5 204:39, 1970a; Pitha, *Biopolymers*, 9:965, 1970b), and morpholino backbones (Summerton, J., et al., U.S. Patent No. 5,142,047, issued 08/25/92; Summerton, J., et al., U.S. Patent No. 5,185,444 issued 02/09/93). A variety of other charged and uncharged polynucleotide analogs have been reported. Numerous backbone modifications are known in the art, including, but not limited to, uncharged linkages (e.g., methyl phosphonates, 10 phosphotriesters, phosphoamidates, and carbamates) and charged linkages (e.g., phosphorothioates and phosphorodithioates).}; and (7) other substances that act as immunostimulating agents to enhance the effectiveness of the VLP immune-stimulating (or vaccine) composition. Alum, CpG oligonucleotides, and MF59 are preferred.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Dosage treatment with the VLP composition may be a single dose schedule or a multiple dose schedule. A multiple dose schedule is one in which a primary course of 20 vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals, chosen to maintain and/or reinforce the immune response, for example at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the subject and be dependent on the judgment of the practitioner.

If prevention of disease is desired, the antigen carrying VLPs are generally administered prior to primary infection with the pathogen of interest. If treatment is desired, e.g., the reduction of symptoms or recurrences, the VLP compositions are generally administered subsequent to primary infection.

30 2.3.2 USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

A number of viral based systems have been developed for use as gene transfer vectors for mammalian host cells. For example, retroviruses (in particular, lentiviral vectors) provide a convenient platform for gene delivery systems. A coding sequence of interest (for example, 35 a sequence useful for gene therapy applications) can be inserted into a gene delivery vector

and packaged in retroviral particles using techniques known in the art. Recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described, including, for example, the following: (U.S. Patent No. 5,219,740; Miller et al. (1989) *BioTechniques* 7:980; Miller, A.D. (1990) *Human Gene Therapy* 1:5; Scarpa et al. (1991) *Virology* 180:849; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033; Boris-Lawrie et al. (1993) *Cur. Opin. Genet. Develop.* 3:102; GB 2200651; EP 0415731; EP 0345242; WO 89/02468; WO 89/05349; WO 89/09271; WO 90/02806; WO 90/07936; WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; WO 93/11230; WO 93/10218; WO 91/02805; in U.S. 5,219,740; U.S. 4,405,712; U.S. 4,861,719; U.S. 4,980,289 and U.S. 4,777,127; in U.S. Serial No. 07/800,921; and in Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53:83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci USA* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

In other embodiments, gene transfer vectors can be constructed to encode a cytokine or other immunomodulatory molecule. For example, nucleic acid sequences encoding native IL-2 and gamma-interferon can be obtained as described in US Patent Nos. 4,738,927 and 5,326,859, respectively, while useful mutants of these proteins can be obtained as described in U.S. Patent No. 4,853,332. Nucleic acid sequences encoding the short and long forms of mCSF can be obtained as described in US Patent Nos. 4,847,201 and 4,879,227, respectively. In particular aspects of the invention, retroviral vectors expressing cytokine or immunomodulatory genes can be produced as described herein (for example, employing the packaging cell lines of the present invention) and in International Application No. PCT US 94/02951, entitled "Compositions and Methods for Cancer Immunotherapy."

Examples of suitable immunomodulatory molecules for use herein include the following: IL-1 and IL-2 (Karupiah et al. (1990) *J. Immunology* 144:290-298, Weber et al. (1987) *J. Exp. Med.* 166:1716-1733, Gansbacher et al. (1990) *J. Exp. Med.* 172:1217-1224, and U.S. Patent No. 4,738,927); IL-3 and IL-4 (Tepper et al. (1989) *Cell* 57:503-512, Golumbek et al. (1991) *Science* 254:713-716, and U.S. Patent No. 5,017,691); IL-5 and IL-6 (Brakenhoff et al. (1987) *J. Immunol.* 139:4116-4121, and International Publication No. WO 90/06370); IL-7 (U.S. Patent No. 4,965,195); IL-8, IL-9, IL-10, IL-11, IL-12, and IL-13 (*Cytokine Bulletin*, Summer 1994); IL-14 and IL-15; alpha interferon (Finter et al. (1991) *Drugs* 42:749-765, U.S. Patent Nos. 4,892,743 and 4,966,843, International Publication No. WO 85/02862, Nagata et al. (1980) *Nature* 284:316-320, Familletti et al. (1981) *Methods in Enz.* 78:387-394, Twu et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2046-2050, and Faktor et

al. (1990) *Oncogene* 5:867-872); beta-interferon (Seif et al. (1991) *J. Virol.* 65:664-671); gamma-interferons (Radford et al. (1991) *The American Society of Hepatology* 20082015, Watanabe et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:9456-9460, Gansbacher et al. (1990) *Cancer Research* 50:7820-7825, Maio et al. (1989) *Can. Immunol. Immunother.* 30:34-42, and U.S. Patent Nos. 4,762,791 and 4,727,138); G-CSF (U.S. Patent Nos. 4,999,291 and 4,810,643); GM-CSF (International Publication No. WO 85/04188).

5 Immunomodulatory factors may also be agonists, antagonists, or ligands for these molecules. For example, soluble forms of receptors can often behave as antagonists for these types of factors, as can mutated forms of the factors themselves.

10 Nucleic acid molecules that encode the above-described substances, as well as other nucleic acid molecules that are advantageous for use within the present invention, may be readily obtained from a variety of sources, including, for example, depositories such as the American Type Culture Collection, or from commercial sources such as British Bio-Technology Limited (Cowley, Oxford England). Representative examples include BBG 12
15 (containing the GM-CSF gene coding for the mature protein of 127 amino acids), BBG 6 (which contains sequences encoding gamma interferon), A.T.C.C. Deposit No. 39656 (which contains sequences encoding TNF), A.T.C.C. Deposit No. 20663 (which contains sequences encoding alpha-interferon), A.T.C.C. Deposit Nos. 31902, 31902 and 39517 (which contain sequences encoding beta-interferon), A.T.C.C. Deposit No. 67024 (which contains a
20 sequence which encodes Interleukin-1b), A.T.C.C. Deposit Nos. 39405, 39452, 39516, 39626 and 39673 (which contain sequences encoding Interleukin-2), A.T.C.C. Deposit Nos. 59399, 59398, and 67326 (which contain sequences encoding Interleukin-3), A.T.C.C. Deposit No. 57592 (which contains sequences encoding Interleukin-4), A.T.C.C. Deposit Nos. 59394 and 59395 (which contain sequences encoding Interleukin-5), and A.T.C.C. Deposit No. 67153
25 (which contains sequences encoding Interleukin-6).

Plasmids containing cytokine genes or immunomodulatory genes (International Publication Nos. WO 94/02951 and WO 96/21015) can be digested with appropriate restriction enzymes, and DNA fragments containing the particular gene of interest can be inserted into a gene transfer vector using standard molecular biology techniques. (See, e.g.,
30 Sambrook et al., *supra*., or Ausbel et al. (eds) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience).

35 Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. For example, plasmids which contain sequences that encode altered cellular products may be

obtained from a depository such as the A.T.C.C., or from commercial sources. Plasmids containing the nucleotide sequences of interest can be digested with appropriate restriction enzymes, and DNA fragments containing the nucleotide sequences can be inserted into a gene transfer vector using standard molecular biology techniques.

5 Alternatively, cDNA sequences for use with the present invention may be obtained from cells which express or contain the sequences, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA. Briefly, mRNA from a cell which expresses the gene of interest can be reverse transcribed with reverse transcriptase using
10 oligo-dT or random primers. The single stranded cDNA may then be amplified by PCR (see U.S. Patent Nos. 4,683,202, 4,683,195 and 4,800,159, see also *PCR Technology: Principles and Applications for DNA Amplification*, Erlich (ed.), Stockton Press, 1989) using oligonucleotide primers complementary to sequences on either side of desired sequences.

15 The nucleotide sequence of interest can also be produced synthetically, rather than cloned, using a DNA synthesizer (e.g., an Applied Biosystems Model 392 DNA Synthesizer, available from ABI, Foster City, California). The nucleotide sequence can be designed with the appropriate codons for the expression product desired. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge (1981) *Nature* 292:756; Nambair et al.
20 (1984) *Science* 223:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311.

The synthetic expression cassettes of the present invention can be employed in the construction of packaging cell lines for use with retroviral vectors.

25 One type of retrovirus, the murine leukemia virus, or "MLV", has been widely utilized for gene therapy applications (see generally Mann et al. (*Cell* 33:153, 1993), Cane and Mulligan (*Proc. Nat'l. Acad. Sci. USA* 81:6349, 1984), and Miller et al., *Human Gene Therapy* 1:5-14, 1990).

30 Lentiviral vectors typically, comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, a promoter operably linked to one or more genes of interest, an origin of second strand DNA synthesis and a 3' lentiviral LTR, wherein the lentiviral vector contains a nuclear transport element. The nuclear transport element may be located either upstream (5') or downstream (3') of a coding sequence of interest (for example, a synthetic Gag or Env expression cassette of the present invention). Within certain embodiments, the nuclear transport element is not RRE. Within one embodiment the packaging signal is an extended packaging signal. Within other embodiments the promoter is a tissue specific promoter, or,

alternatively, a promoter such as CMV. Within other embodiments, the lentiviral vector further comprises an internal ribosome entry site.

A wide variety of lentiviruses may be utilized within the context of the present invention, including for example, lentiviruses selected from the group consisting of HIV, 5 HIV-1, HIV-2, FIV and SIV.

In one embodiment of the present invention synthetic Gag-polymerase expression cassettes are provided comprising a promoter and a sequence encoding synthetic Gag-polymerase and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif.

10 Within yet another aspect of the invention, host cells (e.g., packaging cell lines) are provided which contain any of the expression cassettes described herein. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Gag-polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Gag-polymerase. Packaging cell lines 15 may further comprise a promoter and a sequence encoding tat, rev, or an envelope, wherein the promoter is operably linked to the sequence encoding tat, rev, Env or sequences encoding modified versions of these proteins. The packaging cell line may further comprise a sequence encoding any one or more of nef, vif, vpu or vpr (wild-type or synthetic).

In one embodiment, the expression cassette (carrying, for example, the synthetic Gag-polymerase) is stably integrated. The packaging cell line, upon introduction of a lentiviral 20 vector, typically produces particles. The promoter regulating expression of the synthetic expression cassette may be inducible. Typically, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are essentially free of replication competent virus.

Packaging cell lines are provided comprising an expression cassette which directs the 25 expression of a synthetic *Gag-polymerase* gene or comprising an expression cassette which directs the expression of a synthetic Env genes described herein. (See, also, Andre, S., et al., *Journal of Virology* 72(2):1497-1503, 1998; Haas, J., et al., *Current Biology* 6(3):315-324, 1996) for a description of other modified Env sequences). A lentiviral vector is introduced into the packaging cell line to produce a vector producing cell line.

30 As noted above, lentiviral vectors can be designed to carry or express a selected gene(s) or sequences of interest. Lentiviral vectors may be readily constructed from a wide variety of lentiviruses (see RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985). Representative examples of lentiviruses included HIV, HIV-1, HIV-2, FIV and SIV. Such lentiviruses may either be obtained from patient isolates, or, more

preferably, from depositories or collections such as the American Type Culture Collection, or isolated from known sources using available techniques.

Portions of the lentiviral gene delivery vectors (or vehicles) may be derived from different viruses. For example, in a given recombinant lentiviral vector, LTRs may be derived from an HIV, a packaging signal from SIV, and an origin of second strand synthesis from HrV-2. Lentiviral vector constructs may comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, one or more heterologous sequences, an origin of second strand DNA synthesis and a 3' LTR, wherein said lentiviral vector contains a nuclear transport element that is not RRE.

Briefly, Long Terminal Repeats ("LTRs") are subdivided into three elements, designated U5, R and U3. These elements contain a variety of signals which are responsible for the biological activity of a retrovirus, including for example, promoter and enhancer elements which are located within U3. LTRs may be readily identified in the provirus (integrated DNA form) due to their precise duplication at either end of the genome. As utilized herein, a 5' LTR should be understood to include a 5' promoter element and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector. The 3' LTR should be understood to include a polyadenylation signal, and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector.

The tRNA binding site and origin of second strand DNA synthesis are also important for a retrovirus to be biologically active, and may be readily identified by one of skill in the art. For example, retroviral tRNA binds to a tRNA binding site by Watson-Crick base pairing, and is carried with the retrovirus genome into a viral particle. The tRNA is then utilized as a primer for DNA synthesis by reverse transcriptase. The tRNA binding site may be readily identified based upon its location just downstream from the 5'LTR. Similarly, the origin of second strand DNA synthesis is, as its name implies, important for the second strand DNA synthesis of a retrovirus. This region, which is also referred to as the poly-purine tract, is located just upstream of the 3'LTR.

In addition to a 5' and 3' LTR, tRNA binding site, and origin of second strand DNA synthesis, recombinant retroviral vector constructs may also comprise a packaging signal, as well as one or more genes or coding sequences of interest. In addition, the lentiviral vectors have a nuclear transport element which, in preferred embodiments is not RRE. Representative examples of suitable nuclear transport elements include the element in Rous sarcoma virus (Ogert, et al., *J Virol.* 70, 3834-3843, 1996), the element in Rous sarcoma virus (Liu & Mertz, *Genes & Dev.*, 9, 1766-1789, 1995) and the element in the genome of simian retrovirus type I (Zolotukhin, et al., *J Virol.* 68, 7944-7952, 1994). Other potential elements

include the elements in the histone gene (Kedes, *Annu. Rev. Biochem.* 48, 837-870, 1970), the α -interferon gene (Nagata et al., *Nature* 287, 401-408, 1980), the β -adrenergic receptor gene (Koikka, et al., *Nature* 329, 75-79, 1987), and the c-Jun gene (Hattorie, et al., *Proc. Natl. Acad. Sci. USA* 85, 9148-9152, 1988).

5 Recombinant lentiviral vector constructs typically lack both *Gag-polymerase* and *Env* coding sequences. Recombinant lentiviral vector typically contain less than 20, preferably 15, more preferably 10, and most preferably 8 consecutive nucleotides found in *Gag-polymerase* and *Env* genes. One advantage of the present invention is that the synthetic *Gag-polymerase* expression cassettes, which can be used to construct packaging cell lines for the 10 recombinant retroviral vector constructs, have little homology to wild-type *Gag-polymerase* sequences and thus considerably reduce or eliminate the possibility of homologous recombination between the synthetic and wild-type sequences.

Lentiviral vectors may also include tissue-specific promoters to drive expression of one or more genes or sequences of interest.

15 Lentiviral vector constructs may be generated such that more than one gene of interest is expressed. This may be accomplished through the use of di- or oligo-cistronic cassettes (e.g., where the coding regions are separated by 80 nucleotides or less, *see generally* Levin et al., *Gene* 108:167-174, 1991), or through the use of Internal Ribosome Entry Sites ("IRES").

20 Packaging cell lines suitable for use with the above described recombinant retroviral vector constructs may be readily prepared given the disclosure provided herein. Briefly, the parent cell line from which the packaging cell line is derived can be selected from a variety of mammalian cell lines, including for example, 293, RD, COS-7, CHO, BHK, VERO, HT1080, and myeloma cells.

25 After selection of a suitable host cell for the generation of a packaging cell line, one or more expression cassettes are introduced into the cell line in order to complement or supply in *trans* components of the vector which have been deleted.

30 Representative examples of suitable expression cassettes have been described herein and include synthetic Env, synthetic Gag, synthetic Gag-protease, and synthetic Gag-polymerase expression cassettes, which comprise a promoter and a sequence encoding, e.g., Gag-polymerase and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif. As described above, the native and/or synthetic coding sequences may also be utilized in these expression cassettes.

35 Utilizing the above-described expression cassettes, a wide variety of packaging cell lines can be generated. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Gag-

polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Gag-polymerase. Within other aspects, packaging cell lines are provided comprising a promoter and a sequence encoding tat, rev, Env, or other HIV antigens or epitopes derived therefrom, wherein the promoter is operably linked to the sequence encoding tat, rev, Env, or the HIV antigen or epitope. Within further embodiments, the packaging cell line may comprise a sequence encoding any one or more of nef, vif, vpu or vpr. For example, the packaging cell line may contain only nef, vif, vpu, or vpr alone, nef and vif, nef and vpu, nef and vpr, vif and vpu, vif and vpr, vpu and vpr, nef vif and vpu, nef vif and vpr, nef vpu and vpr, vvir vpu and vpr, or, all four of nef, vif, vpu, and vpr.

In one embodiment, the expression cassette is stably integrated. Within another embodiment, the packaging cell line, upon introduction of a lentiviral vector, produces particles. Within further embodiments the promoter is inducible. Within certain preferred embodiments of the invention, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are free of replication competent virus.

The synthetic cassettes containing modified coding sequences are transfected into a selected cell line. Transfected cells are selected that (i) carry, typically, integrated, stable copies of the HIV coding sequences, and (ii) are expressing acceptable levels of these polypeptides (expression can be evaluated by methods known in the prior art, e.g., see Examples 1-4). The ability of the cell line to produce VLPs may also be verified.

A sequence of interest is constructed into a suitable viral vector as discussed above. This defective virus is then transfected into the packaging cell line. The packaging cell line provides the viral functions necessary for producing virus-like particles into which the defective viral genome, containing the sequence of interest, are packaged. These VLPs are then isolated and can be used, for example, in gene delivery or gene therapy.

Further, such packaging cell lines can also be used to produce VLPs alone, which can, for example, be used as adjuvants for administration with other antigens or in vaccine compositions. Also, co-expression of a selected sequence of interest encoding a polypeptide (for example, an antigen) in the packaging cell line can also result in the entrapment and/or association of the selected polypeptide in/with the VLPs.

Various forms of the different embodiments of the present invention (e.g., constructs) may be combined.

2.4 DNA IMMUNIZATION AND GENE DELIVERY

A variety of HIV polypeptide antigens, particularly Type C HIV antigens, can be used in the practice of the present invention. HIV antigens can be included in DNA immunization

constructs containing, for example, a synthetic Gag expression cassette fused in-frame to a coding sequence for the polypeptide antigen (synthetic or wild-type), where expression of the construct results in VLPs presenting the antigen of interest.

HIV antigens of particular interest to be used in the practice of the present invention include tat, rev, nef, vif, vpu, vpr, and other HIV antigens or epitopes derived therefrom. These antigens may be synthetic (as described herein) or wild-type. Further, the packaging cell line may contain only nef, and HIV-1 (also known as HTLV-III, LAV, ARV, etc.), including, but not limited to, antigens such as gp120, gp41, gp160 (both native and modified); Gag; and pol from a variety of isolates including, but not limited to, HIV_{mb}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2_{UC1} and HIV-2_{UC2}). See, e.g., Myers, et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico; Myers, et al., *Human Retroviruses and Aids, 1990*, Los Alamos, New Mexico: Los Alamos National Laboratory.

To evaluate efficacy, DNA immunization using synthetic expression cassettes of the present invention can be performed, for instance as described in Example 4. Mice are immunized with both the Gag (and/or Env) synthetic expression cassette and the Gag (and/or Env) wild type expression cassette. Mouse immunizations with plasmid-DNAs will show that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the second boost immunization will induce a secondary immune response, for example, after approximately two weeks. Further, the results of CTL assays will show increased potency of synthetic Gag (and/or Env) expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

It is readily apparent that the subject invention can be used to mount an immune response to a wide variety of antigens and hence to treat or prevent a HIV infection, particularly Type C HIV infection.

2.4.1 DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. Furthermore, the desired gene can be isolated directly from cells and tissues containing the same, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al.; *supra*, for a description of techniques used to obtain and

isolate DNA. The gene of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence is assembled from 5 overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature* (1981) 292:756; Nambair et al., *Science* (1984) 223:1299; Jay et al., *J. Biol. Chem.* (1984) 259:6311; Stemmer, W.P.C., (1995) *Gene* 164:49-53.

Next, the gene sequence encoding the desired antigen can be inserted into a vector 10 containing a synthetic expression cassette of the present invention. In certain embodiments, the antigen is inserted into the synthetic Gag coding sequence such that when the combined sequence is expressed it results in the production of VLPs comprising the Gag polypeptide and the antigen of interest, e.g., Env (native or modified) or other antigen(s) (native or modified) derived from HIV. Insertions can be made within the coding sequence or at either 15 end of the coding sequence (5', amino terminus of the expressed Gag polypeptide; or 3', carboxy terminus of the expressed Gag polypeptide)(Wagner, R., et al., *Arch Virol.* 127:117-137, 1992; Wagner, R., et al., *Virology* 200:162-175, 1994; Wu, X., et al., *J. Virol.* 69(6):3389-3398, 1995; Wang, C-T., et al., *Virology* 200:524-534, 1994; Chazal, N., et al., *Virology* 68(1):111-122, 1994; Griffiths, J.C., et al., *J. Virol.* 67(6):3191-3198, 1993; Reicin, 20 A.S., et al., *J. Virol.* 69(2):642-650, 1995).

Up to 50% of the coding sequences of p55Gag can be deleted without affecting the assembly to virus-like particles and expression efficiency (Borsetti, A., et al., *J. Virol.* 72(11):9313-9317, 1998; Gamier, L., et al., *J. Virol.* 72(6):4667-4677, 1998; Zhang, Y., et al., *J. Virol.* 72(3):1782-1789, 1998; Wang, C., et al., *J. Virol.* 72(10): 7950-7959, 1998). In one embodiment of the present invention, immunogenicity of the high level expressing synthetic 25 Gag expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted regions of Gag sequence. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length sequences fused to heterologous 30 polypeptides, can be higher expression/secretion efficiency of the expression product.

When sequences are added to the amino terminal end of Gag, the polynucleotide can contain coding sequences at the 5' end that encode a signal for addition of a myristic moiety to the Gag-containing polypeptide (e.g., sequences that encode Met-Gly).

The ability of Gag-containing polypeptide constructs to form VLPs can be empirically determined following the teachings of the present specification.

The synthetic expression cassettes can also include control elements operably linked to the coding sequence, which allow for the expression of the gene *in vivo* in the subject

5 species. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression.

10 Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook et al., *supra*, as well as a bovine growth hormone terminator sequence.

15 Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence.

20 Furthermore, plasmids can be constructed which include a chimeric antigen-coding gene sequences, encoding, e.g., multiple antigens/epitopes of interest, for example derived from more than one viral isolate.

25 Typically the antigen coding sequences precede or follow the synthetic coding sequence and the chimeric transcription unit will have a single open reading frame encoding both the antigen of interest and the synthetic coding sequences. Alternatively, multi-cistronic cassettes (e.g., bi-cistronic cassettes) can be constructed allowing expression of multiple antigens from a single mRNA using the EMCV IRES, or the like.

Once complete, the constructs are used for nucleic acid immunization using standard 30 gene delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject.

35 A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery

systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; 5 Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109.

A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 10 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

15 Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery. AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Patent Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769 (published 4 March 1993); Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., 20 *Vaccines* 90 (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzyczka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875.

Another vector system useful for delivering the polynucleotides of the present 25 invention is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., et al. (U.S. Patent No. 5,676,950, issued October 14, 1997).

Additional viral vectors which will find use for delivering the nucleic acid molecules encoding the antigens of interest include those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. By way of example, vaccinia virus 30 recombinants expressing the genes can be constructed as follows. The DNA encoding the particular synthetic HIV subtype C polypeptide coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination 35 serves to insert the vaccinia promoter plus the gene encoding the coding sequences of interest

into the viral genome. The resulting TK recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the genes. Recombinant avipox viruses, expressing immunogens from 5 mammalian pathogens, are known to confer protective immunity when administered to non-mammalian species. The use of an avipox vector is particularly desirable in human and other 10 mammalian species since members of the avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as, but not limited to, vectors derived from 15 the Sindbis, Semliki Forest, and Venezuelan Equine Encephalitis viruses, will also find use as viral vectors for delivering the polynucleotides of the present invention (for example, a synthetic Gag-polypeptide encoding expression cassette). For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072; as well as, Dubensky, Jr., T.W., et al., U.S. Patent No. 5,843,723, issued December 1, 1998, and Dubensky, Jr., T.W., U.S. Patent No. 5,789,245, issued August 4, 1998.

A vaccinia based infection/transfection system can be conveniently used to provide 20 for inducible, transient expression of the coding sequences of interest in a host cell. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in 25 the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its 30 translation products. See, e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or 35 to the delivery of genes using other viral vectors, an amplification system can be used that

will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered. Translation of RNA derived from this template will generate T7 RNA polymerase which in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, e.g., International Publication No. WO 94/26911; Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao et al., *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen et al., *Nuc. Acids Res.* (1994) 22:2114-2120; and U.S. Patent No. 5,135,855.

Synthetic expression cassettes of interest can also be delivered without a viral vector. For example, the synthetic expression cassette can be packaged in liposomes prior to delivery to the subject or to cells derived therefrom. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416); mRNA (Malone et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6077-6081); and purified transcription factors (Debs et al., *J. Biol. Chem.* (1990) 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., *Proc. Natl.*

Acad. Sci. USA (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as, from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials.

Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammelar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in *METHODS OF IMMUNOLOGY* (1983), Vol. 101, pp. 512-527; Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483; Wilson et al., *Cell* (1979) 17:77; Deamer and Bangham, *Biochim. Biophys. Acta* (1976) 443:629; Ostro et al., *Biochem. Biophys. Res. Commun.* (1977) 76:836; Fraley et al., *Proc. Natl. Acad. Sci. USA* (1979) 76:3348; Enoch and Strittmatter, *Proc. Natl. Acad. Sci. USA* (1979) 76:145; Fraley et al., *J. Biol. Chem.* (1980) 255:10431; Szoka and Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* (1978) 75:145; and Schaefer-Ridder et al., *Science* (1982) 215:166.

The DNA and/or protein antigen(s) can also be delivered in cochleate lipid compositions similar to those described by Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

The synthetic expression cassette of interest may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected antigen to the immune system and promote trapping and retention of antigens in local lymph nodes. The particles can be phagocytosed by macrophages and can enhance antigen presentation through cytokine release. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Suitable microparticles may also be manufactured in the presence of charged detergents, such as anionic or cationic detergents, to yield microparticles with a surface having a net negative or a net positive charge. For example, microparticles manufactured with anionic detergents, such as

hexadecyltrimethylammonium bromide (CTAB), i.e. CTAB-PLG microparticles, adsorb negatively charged macromolecules, such as DNA. (see, e.g., Int'l Application Number PCT/US99/17308).

Furthermore, other particulate systems and polymers can be used for the *in vivo* or *ex vivo* delivery of the gene of interest. For example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules, are useful for transferring a nucleic acid of interest. Similarly, DEAE dextran-mediated transfection, calcium phosphate precipitation or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like, will find use with the present methods. See, e.g., 5 Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Peptoids (Zuckerman, R.N., et al., U.S. Patent No. 10 5,831,005, issued November 3, 1998) may also be used for delivery of a construct of the present invention.

15 Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering synthetic expression cassettes of the present invention. The particles are coated with the synthetic expression cassette(s) to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful 20 therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744. Also, needle-less injection systems can be used (Davis, H.L., et al, *Vaccine* 12:1503-1509, 1994; Bioject, Inc., Portland, OR).

25 Recombinant vectors carrying a synthetic expression cassette of the present invention are formulated into compositions for delivery to the vertebrate subject. These compositions may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). The compositions will comprise a "therapeutically effective amount" of the gene of interest such that an amount of the antigen can be produced *in vivo* so that an immune response is generated in the individual to which it is administered. The exact amount necessary will vary depending on the subject being treated; the age and general condition of 30 the subject to be treated; the capacity of the subject's immune system to synthesize antibodies; the degree of protection desired; the severity of the condition being treated; the particular antigen selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. Thus, a "therapeutically effective amount" will fall in a relatively broad range that can be determined 35 through routine trials.

The compositions will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, polyethyleneglycol, hyaluronic acid, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Certain facilitators of nucleic acid uptake and/or expression can also be included in the compositions or coadministered, such as, but not limited to, bupivacaine, cardiotoxin and sucrose.

Once formulated, the compositions of the invention can be administered directly to the subject (e.g., as described above) or, alternatively, delivered *ex vivo*, to cells derived from the subject, using methods such as those described above. For example, methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and can include, e.g., dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, lipofectamine and LT-1 mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) (with or without the corresponding antigen) in liposomes, and direct microinjection of the DNA into nuclei.

Direct delivery of synthetic expression cassette compositions *in vivo* will generally be accomplished with or without viral vectors, as described above, by injection using either a conventional syringe or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). The constructs can be injected either subcutaneously, epidermally, intradermally, intramucosally such as nasally, rectally and vaginally, intraperitoneally, intravenously, orally or intramuscularly. Delivery of DNA into cells of the epidermis is particularly preferred as this mode of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Other modes of administration include oral and pulmonary administration, suppositories, needle-less injection, transcutaneous and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. Administration of nucleic acids may also be combined with administration of peptides or other substances.

2.4.2 EX VIVO DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION

In one embodiment, T cells, and related cell types (including but not limited to antigen presenting cells, such as, macrophage, monocytes, lymphoid cells, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), can be used for *ex vivo* delivery of the synthetic expression cassettes of the present invention. T cells can be isolated from peripheral blood lymphocytes (PBLs) by a variety of procedures known to those skilled in the art. For example, T cell populations can be "enriched" from a population of PBLs through

the removal of accessory and B cells. In particular, T cell enrichment can be accomplished by the elimination of non-T cells using anti-MHC class II monoclonal antibodies. Similarly, other antibodies can be used to deplete specific populations of non-T cells. For example, anti-Ig antibody molecules can be used to deplete B cells and anti-MacI antibody molecules 5 can be used to deplete macrophages.

T cells can be further fractionated into a number of different subpopulations by techniques known to those skilled in the art. Two major subpopulations can be isolated based on their differential expression of the cell surface markers CD4 and CD8. For example, following the enrichment of T cells as described above, CD4⁺ cells can be enriched using 10 antibodies specific for CD4 (see Coligan et al., *supra*). The antibodies may be coupled to a solid support such as magnetic beads. Conversely, CD8⁺ cells can be enriched through the use of antibodies specific for CD4 (to remove CD4⁺ cells), or can be isolated by the use of CD8 antibodies coupled to a solid support. CD4 lymphocytes from HIV-1 infected patients can be expanded *ex vivo*, before or after transduction as described by Wilson et. al. (1995) *J.*

15 *Infect. Dis.* 172:88.

Following purification of T cells, a variety of methods of genetic modification known to those skilled in the art can be performed using non-viral or viral-based gene transfer vectors constructed as described herein. For example, one such approach involves 20 transduction of the purified T cell population with vector-containing supernatant of cultures derived from vector producing cells. A second approach involves co-cultivation of an irradiated monolayer of vector-producing cells with the purified T cells. A third approach involves a similar co-cultivation approach; however, the purified T cells are pre-stimulated with various cytokines and cultured 48 hours prior to the co-cultivation with the irradiated vector producing cells. Pre-stimulation prior to such transduction increases effective gene 25 transfer (Nolta et al. (1992) *Exp. Hematol.* 20:1065). Stimulation of these cultures to proliferate also provides increased cell populations for re-infusion into the patient. Subsequent to co-cultivation, T cells are collected from the vector producing cell monolayer, expanded, and frozen in liquid nitrogen.

Gene transfer vectors, containing one or more synthetic expression cassette of the 30 present invention (associated with appropriate control elements for delivery to the isolated T cells) can be assembled using known methods.

Selectable markers can also be used in the construction of gene transfer vectors. For example, a marker can be used which imparts to a mammalian cell transduced with the gene transfer vector resistance to a cytotoxic agent. The cytotoxic agent can be, but is not limited 35 to, neomycin, aminoglycoside, tetracycline, chloramphenicol, sulfonamide, actinomycin,

netropsin, distamycin A, anthracycline, or pyrazinamide. For example, neomycin phosphotransferase II imparts resistance to the neomycin analogue geneticin (G418).

The T cells can also be maintained in a medium containing at least one type of growth factor prior to being selected. A variety of growth factors are known in the art which sustain the growth of a particular cell type. Examples of such growth factors are cytokine mitogens such as rIL-2, IL-10, IL-12, and IL-15, which promote growth and activation of lymphocytes. Certain types of cells are stimulated by other growth factors such as hormones, including human chorionic gonadotropin (hCG) and human growth hormone. The selection of an appropriate growth factor for a particular cell population is readily accomplished by one of skill in the art.

For example, white blood cells such as differentiated progenitor and stem cells are stimulated by a variety of growth factors. More particularly, IL-3, IL-4, IL-5, IL-6, IL-9, GM-CSF, M-CSF, and G-CSF, produced by activated T_H and activated macrophages, stimulate myeloid stem cells, which then differentiate into pluripotent stem cells, granulocyte-monocyte progenitors, eosinophil progenitors, basophil progenitors, megakaryocytes, and erythroid progenitors. Differentiation is modulated by growth factors such as GM-CSF, IL-3, IL-6, IL-11, and EPO.

Pluripotent stem cells then differentiate into lymphoid stem cells, bone marrow stromal cells, T cell progenitors, B cell progenitors, thymocytes, T_H Cells, T_C cells, and B cells. This differentiation is modulated by growth factors such as IL-3, IL-4, IL-6, IL-7, GM-CSF, M-CSF, G-CSF, IL-2, and IL-5.

Granulocyte-monocyte progenitors differentiate to monocytes, macrophages, and neutrophils. Such differentiation is modulated by the growth factors GM-CSF, M-CSF, and IL-8. Eosinophil progenitors differentiate into eosinophils. This process is modulated by GM-CSF and IL-5.

The differentiation of basophil progenitors into mast cells and basophils is modulated by GM-CSF, IL-4, and IL-9. Megakaryocytes produce platelets in response to GM-CSF, EPO, and IL-6. Erythroid progenitor cells differentiate into red blood cells in response to EPO.

Thus, during activation by the CD3-binding agent, T cells can also be contacted with a mitogen, for example a cytokine such as IL-2. In particularly preferred embodiments, the IL-2 is added to the population of T cells at a concentration of about 50 to 100 $\mu\text{g}/\text{ml}$. Activation with the CD3-binding agent can be carried out for 2 to 4 days.

Once suitably activated, the T cells are genetically modified by contacting the same with a suitable gene transfer vector under conditions that allow for transfection of the vectors

into the T cells. Genetic modification is carried out when the cell density of the T cell population is between about 0.1×10^6 and 5×10^6 , preferably between about 0.5×10^6 and 2×10^6 . A number of suitable viral and nonviral-based gene transfer vectors have been described for use herein.

After transduction, transduced cells are selected away from non-transduced cells using known techniques. For example, if the gene transfer vector used in the transduction includes a selectable marker which confers resistance to a cytotoxic agent, the cells can be contacted with the appropriate cytotoxic agent, whereby non-transduced cells can be negatively selected away from the transduced cells. If the selectable marker is a cell surface marker, the cells can be contacted with a binding agent specific for the particular cell surface marker, whereby the transduced cells can be positively selected away from the population. The selection step can also entail fluorescence-activated cell sorting (FACS) techniques, such as where FACS is used to select cells from the population containing a particular surface marker, or the selection step can entail the use of magnetically responsive particles as retrievable supports for target cell capture and/or background removal.

More particularly, positive selection of the transduced cells can be performed using a FACS cell sorter (e.g. a FACSVantageTM Cell Sorter, Becton Dickinson Immunocytometry Systems, San Jose, CA) to sort and collect transduced cells expressing a selectable cell surface marker. Following transduction, the cells are stained with fluorescent-labeled antibody molecules directed against the particular cell surface marker. The amount of bound antibody on each cell can be measured by passing droplets containing the cells through the cell sorter. By imparting an electromagnetic charge to droplets containing the stained cells, the transduced cells can be separated from other cells. The positively selected cells are then harvested in sterile collection vessels. These cell sorting procedures are described in detail, for example, in the FACSVantageTM Training Manual, with particular reference to sections 3-11 to 3-28 and 10-1 to 10-17.

Positive selection of the transduced cells can also be performed using magnetic separation of cells based on expression of a particular cell surface marker. In such separation techniques, cells to be positively selected are first contacted with specific binding agent (e.g., an antibody or reagent that interacts specifically with the cell surface marker). The cells are then contacted with retrievable particles (e.g., magnetically responsive particles) which are coupled with a reagent that binds the specific binding agent (that has bound to the positive cells). The cell-binding agent-particle complex can then be physically separated from non-labeled cells, for example using a magnetic field. When using magnetically responsive particles, the labeled cells can be retained in a container using a magnetic field while the

negative cells are removed. These and similar separation procedures are known to those of ordinary skill in the art.

Expression of the vector in the selected transduced cells can be assessed by a number of assays known to those skilled in the art. For example, Western blot or Northern analysis can be employed depending on the nature of the inserted nucleotide sequence of interest.

Once expression has been established and the transformed T cells have been tested for the presence of the selected synthetic expression cassette, they are ready for infusion into a patient via the peripheral blood stream.

The invention includes a kit for genetic modification of an *ex vivo* population of primary mammalian cells. The kit typically contains a gene transfer vector coding for at least one selectable marker and at least one synthetic expression cassette contained in one or more containers, ancillary reagents or hardware, and instructions for use of the kit.

2.4.3 FURTHER DELIVERY REGIMES

Any of the polynucleotides (*e.g.*, expression cassettes) or polypeptides described herein (delivered by any of the methods described above) can also be used in combination with other DNA delivery systems and/or protein delivery systems. Non-limiting examples include co-administration of these molecules, for example, in prime-boost methods where one or more molecules are delivered in a “priming” step and, subsequently, one or more molecules are delivered in a “boosting” step. In certain embodiments, the delivery of one or more nucleic acid-containing compositions and is followed by delivery of one or more nucleic acid-containing compositions and/or one or more polypeptide-containing compositions (*e.g.*, polypeptides comprising HIV antigens). In other embodiments, multiple nucleic acid “primes” (of the same or different nucleic acid molecules) can be followed by multiple polypeptide “boosts” (of the same or different polypeptides). Other examples include multiple nucleic acid administrations and multiple polypeptide administrations.

In any method involving co-administration, the various compositions can be delivered in any order. Thus, in embodiments including delivery of multiple different compositions or molecules, the nucleic acids need not be all delivered before the polypeptides. For example, the priming step may include delivery of one or more polypeptides and the boosting comprises delivery of one or more nucleic acids and/or one more polypeptides. Multiple polypeptide administrations can be followed by multiple nucleic acid administrations or polypeptide and nucleic acid administrations can be performed in any order. In any of the embodiments described herein, the nucleic acid molecules can encode all, some or none of the polypeptides. Thus, one or more of the nucleic

acid molecules (e.g., expression cassettes) described herein and/or one or more of the polypeptides described herein can be co-administered in any order and via any administration routes. Therefore, any combination of polynucleotides and/or polypeptides described herein can be used to generate elicit an immune reaction.

5 **EXPERIMENTAL**

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

Example 1

Generation of Synthetic Expression Cassettes

15 **A. Modification of HIV-1 Env, Gag, Pol Nucleic Acid Coding Sequences**

The Pol coding sequences were selected from Type C strain AF110975. The Gag coding sequences were selected from the Type C strains AF110965 and AF110967. The Env coding sequences were selected from Type C strains AF110968 and AF110975. These sequences were manipulated to maximize expression of their gene products.

20 First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed 25 human codons prefer the nucleotides G or C. The coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences (Schneider R, et al., *J Virol.* 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements are inactivated by introducing multiple point mutations that do not alter the reading frame of the encoded proteins. Figures 5 and 6 (SEQ ID Nos: 3, 4, 20 and 21) show the location of some remaining INS in synthetic sequences derived from strains AF110965 and AF110967. The changes made to these sequences are boxed in the Figures. In Figures 5 and

6, the top line depicts a modified sequence of Gag polypeptides from the indicated strains. The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to further remove INS. Thus, when the changes indicated in the boxed regions are made, the resulting sequences correspond to the sequences depicted in Figures 1 and 2, respectively.

5 The synthetic coding sequences are assembled by methods known in the art, for example by companies such as the Midland Certified Reagent Company (Midland, Texas).

In one embodiment of the invention, sequences encoding Pol-polypeptides are included with the synthetic Gag or Env sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, modified Gag/Env expression cassette.

10 Because synthetic HIV-1 Pol expresses the functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it may be helpful in some instances to inactivate RT and INT functions. Several deletions or mutations in the RT and INT coding regions can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York. ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* 9(1):58-68; Wakefield, J. K., et al., (1992) *Journal Of Virology* 66(11):6806-6812; Esnouf, R., et al., (1995) *Nature Structural Biology* 2(4):303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* 282(2):359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* 73 (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 90(13):6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* 269(46):29279-29287; Goldgur, Y., et al., (1998) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 95(16):9150-9154; Goette, M., et al., (1998) *Journal Of Biological Chemistry* 273(17):10139-10146; Gorton, J. L., et al., (1998) *Journal of Virology* 72(6):5046-5055; Engelman, A., et al., (1997) *Journal Of Virology* 71(5):3507-3514; Dyda, F., et al., *Science* 266(5193):1981-1986; Davies, J. F., et al., (1991) *Science* 252(5002):88-95; Bujacz, G., et al., (1996) *Fefs Letters* 398(2-3):175-178; Beard, W. A., et al., (1996) *Journal Of Biological Chemistry* 271(21):12213-12220; Kohlstaedt, L. A., et al., (1992) *Science* 256(5065):1783-30 1790; Krug, M. S. and Berger, S. L. (1991) *Biochemistry* 30(44):10614-10623; Mazumder, A., et al., (1996) *Molecular Pharmacology* 49(4):621-628; Palaniappan, C., et al., (1997) *Journal Of Biological Chemistry* 272(17):11157-11164; Rodgers, D. W., et al., (1995) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 92(4):1222-1226; Sheng, N. and Dennis, D. (1993) *Biochemistry* 32(18):4938-4942; Spence, R. A., et al., (1995) *Science* 267(5200):988-993.}

Furthermore selected B- and/or T-cell epitopes can be added to the Pol constructs (e.g., 3' of the truncated INT or within the deletions of the RT- and INT-coding sequence) to replace and augment any epitopes deleted by the functional modifications of RT and INT. Alternately, selected B- and T-cell epitopes (including CTL epitopes) from RT and INT can 5 be included in a minimal VLP formed by expression of the synthetic Gag or synthetic Pol cassette, described above. (For descriptions of known HIV B- and T-cell epitopes see, HIV Molecular Immunology Database CTL Search Interface; Los Alamos Sequence Compendia, 1987-1997; Internet address: <http://hiv-web.lanl.gov/immunology/index.html>.)

The resulting modified coding sequences are presented as a synthetic Env expression 10 cassette; a synthetic Gag expression cassette; a synthetic Pol expression cassette. A common Gag region (Gag-common) extends from nucleotide position 844 to position 903 (SEQ ID NO:1), relative to AF110965 (or from approximately amino acid residues 282 to 301 of SEQ ID NO:17) and from nucleotide position 841 to position 900 (SEQ ID NO:2), relative to AF110967 (or from approximately amino acid residues 281 to 300 of SEQ ID NO:22). A 15 common Env region (Env-common) extends from nucleotide position 1213 to position 1353 (SEQ ID NO:5) and amino acid positions 405 to 451 of SEQ ID NO:23, relative to AF110968 and from nucleotide position 1210 to position 1353 (SEQ ID NO:11) and amino acid positions 404-451 (SEQ ID NO:24), relative to AF110975.

The synthetic DNA fragments for Pol, Gag and Env are cloned into the following 20 eucaryotic expression vectors: pCMVKm2, for transient expression assays and DNA immunization studies, the pCMVKm2 vector is derived from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) and comprises a kanamycin selectable marker, a ColE1 origin of replication, a CMV promoter enhancer and Intron A, followed by an insertion site 25 for the synthetic sequences described below followed by a polyadenylation signal derived from bovine growth hormone -- the pCMVKm2 vector differs from the pCMV-link vector only in that a polylinker site is inserted into pCMVKm2 to generate pCMV-link; pESN2dhfr and pCMVPLEdhfr, for expression in Chinese Hamster Ovary (CHO) cells; and, pAcC13, a shuttle vector for use in the Baculovirus expression system (pAcC13, is derived from pAcC12 which is described by Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990).

30 Briefly, construction of pCMVPLEdhfr was as follows.

To construct a DHFR cassette, the EMCV IRES (internal ribosome entry site) leader was PCR-amplified from pCite-4a+ (Novagen, Inc., Milwaukee, WI) and inserted into pET-23d (Novagen, Inc., Milwaukee, WI) as an *Xba-Nco* fragment to give pET-EMCV. The *dhfr* gene was PCR-amplified from pESN2dhfr to give a product with a Gly-Gly-Gly-Ser spacer in 35 place of the translation stop codon and inserted as an *Nco-Bam*H1 fragment to give pET-E-

DHFR. Next, the attenuated *neo* gene was PCR amplified from a pSV2Neo (Clontech, Palo Alto, CA) derivative and inserted into the unique *Bam*H1 site of pET-E-DHFR to give pET-E-DHFR/*Neo*_(m2). Finally the bovine growth hormone terminator from pCDNA3 (Invitrogen, Inc., Carlsbad, CA) was inserted downstream of the *neo* gene to give pET-E-

5 DHFR/*Neo*_(m2)BGHt. The EMCV-*dhfr/neo* selectable marker cassette fragment was prepared by cleavage of pET-E-DHFR/*Neo*_(m2)BGHt.

The CMV enhancer/promoter plus Intron A was transferred from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) as a *Hind*III-*Sal*I fragment into pUC19 (New England Biolabs, Inc., Beverly, MA). The vector backbone of pUC19 was deleted from the 10 NdeI to the SapI sites. The above described DHFR cassette was added to the construct such that the EMCV IRES followed the CMV promoter. The vector also contained an amp' gene and an SV40 origin of replication.

B. Defining of the Major Homology Region (MHR) of HIV-1 p55Gag

15 The Major Homology Region (MHR) of HIV-1 p55 (Gag) is located in the p24-CA sequence of Gag. It is a conserved stretch of approximately 20 amino acids. The position in the wild type AF110965 Gag protein is from 282-301 (SEQ ID NO:25) and spans a region from 844-903 (SEQ ID NO:26) for the Gag DNA-sequence. The position in the synthetic Gag protein is also from 282-301 (SEQ ID NO:25) and spans a region from 844-903 (SEQ ID 20 NO:1) for the synthetic Gag DNA-sequence. The position in the wild type and synthetic AF110967 Gag protein is from 281-300 (SEQ ID NO:27) and spans a region from 841-900 (SEQ ID NO:2) for the modified Gag DNA-sequence. Mutations or deletions in the MHR can severely impair particle production (Borsetti, A., et al., *J. Virol.* 72(11):9313-9317, 1998; Mammano, F., et al., *J Virol* 68(8):4927-4936, 1994).

25 Percent identity to this sequence can be determined, for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5.

C. Defining of the Common Sequence Region of HIV-1 Env

30 The common sequence region (CSR) of HIV-1 Env is located in the C4 sequence of Env. It is a conserved stretch of approximately 47 amino acids. The position in the wild type and synthetic AF110968 Env protein is from approximately amino acid residue 405 to 451 (SEQ ID NO:28) and spans a region from 1213 to 1353 (SEQ ID NO:5) for the Env DNA-sequence. The position in the wild type and synthetic AF110975 Env protein is

from approximately amino acid residue 404 to 451 (SEQ ID NO:29) and spans a region from 1210 to 1353 (SEQ ID NO:11) for the Env DNA-sequence.

Percent identity to this sequence can be determined, for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5.

Various forms of the different embodiments of the invention, described herein, may be combined.

D. Exemplary HIV Sequences Derived from South African HIV Type C Strains

HIV coding sequences of novel Type C isolates were obtained. Polypeptide-coding sequences were manipulated to maximize expression of their gene products.

As described above, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Shown below in Table C are exemplary wild-type and synthetic sequences derived from a novel South African HIV Type C isolate, clone 8_5_TV1_C.ZA. Table D shows exemplary synthetic Env sequences derived from a novel South African HIV Type C isolate, clone 8_2_TV1_C.ZA. Table E shows wild-type and synthetic sequences derived from South African HIV Type C strain 12-5_1_TV2_C.ZA.

Table C

Name	SEQ ID	Description
C4_Env_TV1_C_ZA_opt short	46	synthetic sequence of short Env "common region"
5 C4_Env_TV1_C_ZA_opt	47	synthetic sequence of Env "common region"
C4_Env_TV1_C_ZA_wt	48	wild type 8_5_TV1_C.ZA Env sequence
Envgp160_TV1_C_ZAopt	49	synthetic Env gp160
Envgp160_TV1_C_ZAwt	50	wild type 8_5_TV1_C.ZA Env gp160 sequence
Gag_TV1_C_ZAopt	51	synthetic sequence of Gag
10 Gag_TV1_C_ZAwt	52	wild type 8_5_TV1_C.ZA Gag sequence
Gag_TV1_ZA_MHROpt	53	synthetic sequence of Gag major homology region
Gag_TV1_ZA_MHRwt	54	wild type 8_5_TV1_C.ZA Gag major homology region sequence
Nef_TV1_C_ZAopt	55	synthetic sequence of Nef
Nef_TV1_C_ZAwt	56	wild type 8_5_TV1_C.ZA Nef sequence
15 NefD125G_TV1_C_ZAopt	57	synthetic sequence of Nef, including mutation at position 125 resulting in non-functional gene product
p15RNaseH_TV1_C_ZAopt	58	synthetic sequence of RNaseH (p15 of Pol)
p15RNaseH_TV1_C_ZAwt	59	wild type 8_5_TV1_C.ZA RNaseH sequence
p31Int_TV1_C_ZAopt	60	synthetic sequence of Integrase (p31 of Pol)
p31Int_TV1_C_ZAwt	61	wild type 8_5_TV1_C.ZA Integrase sequence
20 Pol_TV1_C_ZAopt	62	synthetic sequence of Pol
Pol_TV1_C_ZAwt	63	wild type 8_5_TV1_C.ZA Pol sequence
Prot_TV1_C_ZAopt	64	synthetic sequence of Prot
Prot_TV1_C_ZAwt	65	wild type 8_5_TV1_C.ZA Prot sequence
Protina_TV1_C_ZAopt	66	synthetic sequence of Prot including mutation resulting in inactivation of protease

	Protina_TV1_C_ZAwt	67	wild type 8_5_TV1_C.ZA Prot sequence, including mutation resulting in inactivation of protease.
5	ProtinaRTmut_TV1_C_ZAo p	68	synthetic sequence of Prot and reverse transcriptase (RT), including mutation resulting in inactivation of protease and mutation resulting in inactivation of RT.
10	ProtinaRTmut_TV1_C_ZA wt	69	wild type 8_5_TV1_C.ZA Prot and RT, mutation resulting in inactivation of protease and mutation resulting in inactivation of RT.
15	ProtwtRTwt_TV1_C_ZAopt	70	synthetic sequences of Prot and RT
20	ProtwtRTwt_TV1_C_ZAwt	71	wild type 8_5_TV1_C.ZA Prot and RT
	RevExon1_TV1_C_ZAopt	72	synthetic sequence of exon 1 of Rev
	RevExon1_TV1_C_ZAwt	73	wild type 8_5_TV1_C.ZA of exon 1 of Rev
	RevExon2_TV1_C_ZAopt-2	74	synthetic sequence of exon 2 of Rev
	RevExon2_TV1_C_ZAwt	75	wild type 8_5_TV1_C.ZA of exon 2 of Rev
	RT_TV1_C_ZAopt	76	synthetic sequence of RT
	RT_TV1_C_ZAwt	77	wild type 8_5_TV1_C.ZA RT
	RTmut_TV1_C_ZAopt	78	synthetic sequence of RT, including mutation resulting in inactivation of RT
15	RTmut_TV1_C_ZAwt	79	wild type 8_5_TV1_C.ZA RT, including mutation resulting in inactivation of RT
20	TatC22Exon1_TV1_C_ZAo pt	80	synthetic sequence of exon 1 of Tat, including mutation resulting in non-functional Tat gene product
	TatExon1_TV1_C_ZAopt	81	synthetic sequence of exon 1 of Tat
	TatExon1_TV1_C_ZAwt	82	wild type 8_5_TV1_C.ZA exon 1 of Tat
	TatExon2_TV1_C_ZAopt	83	synthetic sequence of exon 2 of Tat
	TatExon2_TV1_C_ZAwt	84	wild type 8_5_TV1_C.ZA exon 2 of Tat
	Vif_TV1_C_ZAopt	85	synthetic sequence of Vif
	Vif_TV1_C_ZAwt	86	wild type 8_5_TV1_C.ZA Vif
	Vpr_TV1_C_ZAopt	87	synthetic sequence of Vpr

	Vpr_TV1_C_ZAwt	88	wild type 8_5_TV1_C.ZA Vpr
	Vpu_TV1_C_ZAopt	89	synthetic sequence of Vpu
	Vpu_TV1_C_ZAwt	90	wild type 8_5_TV1_C.ZA Vpu
5	revexon1_2 TV1 C ZAopt	91	synthetic sequence of exons 1 and 2 of Rev
	RevExon1_2_TV1_C_ZAwt	92	wild type 8_5_TV1_C.ZA Rev (exons 1 and 2)
	TatC22Exon1_2_TV1_C_ZAopt	93	synthetic sequence of exons 1 and 2 of Tat, including mutation in exon 1 resulting in non-functional Tat gene product
	TatExon1_2_TV1_C_ZAopt	94	synthetic sequence of exons 1 and 2 of Tat
	TatExon1_2_TV1_C_ZAwt	95	wild type 8_5_TV1_C.ZA Tat (exons 1 and 2)
10.	NefD125G-Myr_TV1_C_ZAopt	96	synthetic sequence of Nef, including mutation eliminating myristylation site.

Table D

Name	Seq Id	Description
gp120mod.TV1.delV2	119	synthetic sequence of Env gp120, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.delV2	120	synthetic sequence of Env gp140, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
5 gp140mod.TV1.mut7.delV2	121	synthetic sequence of Env gp140, including V2 deletion and mutation in cleavage site and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.delV1V2	122	synthetic sequence of Env gp160, including V1/V2 deletion and modified leader derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.delV2	123	synthetic sequence of Env gp160, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
10 gp160mod.TV1.mut7.delV2	124	synthetic sequence of Env gp160, including V2 deletion; a mutation in cleavage site; and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.tpa1	125	synthetic sequence of Env gp160, TPA1 leader
gp160mod.TV1	126	synthetic sequence of Env gp160, including modified leader sequences derived from wild-type (8_2_TV1_C.ZA) sequences
gp160mod.TV1.wtLnative	127	synthetic sequence of Env gp160, including wild type 8_2_TV1_C.ZA (unmodified) leader
gp140.mod.TV1.tpa1	131	synthetic sequence of Env gp140, TPA1 leader
15 gp140mod.TV1	132	synthetic sequence of Env gp140, including modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.wtLnative	133	synthetic sequence of Env gp120, including wild type 8_2_TV1_C.ZA (unmodified) leader sequence.

As noted above, Env-encoding constructs can be prepared using any of the full-length of gp160 constructs. For example, a gp140 form (SEQ ID NO:132) was made by truncating gp160 (SEQ ID NO:126) at nucleotide 2064; gp120 was made by truncating gp160 (SEQ ID

NO:126) at nucleotide 1551 (SEQ ID NO:126). Additional gp140 and gp120 forms can be made using the methods described herein. One or more stop codons are typically added (e.g., nucleotides 2608 to 2610 of SEQ ID NO:126). Further, the wild-type leader sequence can be modified and/or replaced with other leader sequences (e.g., TPA1 leader sequences).

5 Thus, the polypeptide gp160 includes the coding sequences for gp120 and gp41. The polypeptide gp41 is comprised of several domains including an oligomerization domain (OD) and a transmembrane spanning domain (TM). In the native envelope, the oligomerization domain is required for the non-covalent association of three gp41 polypeptides to form a trimeric structure: through non-covalent interactions with the gp41 trimer (and itself), the
10 gp120 polypeptides are also organized in a trimeric structure. A cleavage site (or cleavage sites) exists approximately between the polypeptide sequences for gp120 and the polypeptide sequences corresponding to gp41. This cleavage site(s) can be mutated to prevent cleavage at the site. The resulting gp140 polypeptide corresponds to a truncated form of gp160 where the transmembrane spanning domain of gp41 has been deleted. This gp140 polypeptide can exist
15 in both monomeric and oligomeric (*i.e.* trimeric) forms by virtue of the presence of the oligomerization domain in the gp41 moiety. In the situation where the cleavage site has been mutated to prevent cleavage and the transmembrane portion of gp41 has been deleted the resulting polypeptide product is designated "mutated" gp140 (e.g., gp140.mut). As will be apparent to those in the field, the cleavage site can be mutated in a variety of ways. In the
20 exemplary constructs described herein (e.g., SEQ ID NO:121 and SEQ ID NO:124), the mutation in the gp120/gp41 cleavage site changes the wild-type amino acid sequence KRRVVQREKR (SEQ ID NO:129) to ISSVHQSEKS (SEQ ID NO:130).

In yet other embodiments, hypervariable region(s) were deleted, N-glycosylation sites were removed and/or cleavage sites mutated. Exemplary constructs having variable region deletions (V1 and/or V2), V2 deletes were constructed by deleting nucleotides from approximately 499 to approximately 593 (relative to SEQ ID NO:128) and V1/V2 deletes were constructed by deleting nucleotides from approximately 375 to approximately 602 (relative to SEQ ID NO:128). The relative locations of V1 and/or V2 regions can also be readily determined by alignment to the regions shown in Table A. Table E shows wild-type
25 and synthetic sequences derived from South African HIV Type C strain 12-5_1_TV2_C.ZA.

Table E

Name	SEQ ID	Description
Envgp160_TV2_C_ZAopt	97	synthetic sequence of Env gp160

	Envgp160_TV2_C_ZAwt	98	wild type 12-5_1_TV2_C.ZA Env gp160.
	Gag_TV2_C_ZAopt	99	synthetic sequence of Gag
	Gag_TV2_C_ZAwt	100	wild type 12-5_1_TV2_C.ZA Gag
	Nef_TV2_C_ZAopt	101	synthetic sequence of Nef
5	Nef_TV2_C_ZAwt	102	wild type 12-5_1_TV2_C.ZA Nef
	Pol_TV2_C_ZAopt	103	synthetic sequence of Pol
	Pol_TV2_C_ZAwt	104	wild type 12-5_1_TV2_C.ZA of Pol
	RevExon1_TV2_C_ZAopt	105	synthetic sequence of exon 1 of Rev
	RevExon1_TV2_C_ZAwt	106	wild type 12-5_1_TV2_C.ZA of exon 1 of Rev
10	RevExon2_TV2_C_ZAopt	107	synthetic sequence of exon 2 of Rev
	RevExon2_TV2_C_ZAwt	108	wild type 12-5_1_TV2_C.ZA of exon 2 of Rev
	TatExon1_TV2_C_ZAopt	109	synthetic sequence of exon 1 of Tat
	TatExon1_TV2_C_ZAwt	110	wild type 12-5_1_TV2_C.ZA of exon 1 of Tat
	TatExon2_TV2_C_ZAopt	111	synthetic sequence of exon 2 of Tat
15	TatExon2_TV2_C_ZAwt	112	wild type 12-5_1_TV2_C.ZA of exon 2 of Tat
	Vif_TV2_C_ZAopt	113	synthetic sequence of Vif
	Vif_TV2_C_ZAwt	114	wild type 12-5_1_TV2_C.ZA of Vif
	Vpr_TV2_C_ZAopt	115	synthetic sequence of Vpr
	Vpr_TV2_C_ZAwt	116	wild type 12-5_1_TV2_C.ZA of Vpr
20	Vpu_TV2_C_ZAopt	117	synthetic sequence of Vpu
	Vpu_TV2_C_ZAwt	118	wild type 12-5_1_TV2_C.ZA of Vpu

It will be readily apparent that sequences derived from any HIV type C strain or clone can be modified as described herein in order to achieve desirable modifications in that strain.

25 Additionally, polyproteins can be constructed by fusing in-frame two or more polynucleotide sequences encoding polypeptide or peptide products. Further, polycistronic coding sequences may be produced by placing two or more polynucleotide sequences encoding polypeptide products adjacent each other, typically under the control of one promoter, wherein each polypeptide coding sequence may be modified to include sequences for internal ribosome binding sites.

The sequences of the present invention, for example, the modified (synthetic) polynucleotide sequences encoding HIV polypeptides, may be modified by deletions, point mutations, substitutions, frame-shifts, and/or further genetic modifications (for example, mutations leading to inactivation of an activity associated with a polypeptide, e.g., mutations that inactivate protease, tat, or reverse transcriptase activity). Such modifications are taught generally in the art and may be applied in the context of the teachings of the present invention. For example, sites corresponding to the "Regions of the HIV Genome" listed in Table A may be modified in the corresponding regions of the novel sequences disclosed herein in order to achieve desirable modifications. Further, the modified (synthetic) polynucleotide sequences of the present invention can be combined for use, e.g., in an composition for generating an immune response in a subject, in a variety of ways, including but not limited to the following ways: multiple individual expression cassettes each comprising one polynucleotide sequence of the present invention (e.g., a gag-expression cassette, an env expression cassette, and a rev expression cassette, or a pol-expression cassette, a vif expression cassette, and a vpr expression cassette, etc.); polyproteins produced by in-frame fusions of multiple polynucleotides of the present invention, and polycistronic polynucleotides produced using multiple polynucleotides of the present invention.

Example 2

Expression Assays for the Synthetic Coding Sequences

A. Type C HIV Coding Sequences

The wild-type Subtype C HIV coding (for example from AF110965, AF110967, AF110968, AF110975, as well as novel South African strains 8_5_TV1_C.ZA, 8_2_TV1_C.ZA and 12-5_1_TV2_C.ZA) sequences are cloned into expression vectors having the same features as the vectors into which the synthetic sequences are cloned.

Expression efficiencies for various vectors carrying the wild-type and synthetic sequences are evaluated as follows. Cells from several mammalian cell lines (293, RD, COS-7, and CHO; all obtained from the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209) are transfected with 2 µg of DNA in transfection reagent LT1 (PanVera Corporation, 545 Science Dr., Madison, WI). The cells are incubated for 5 hours in reduced serum medium (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium is then replaced with normal medium as follows: 293 cells, IMDM, 10% fetal calf

serum, 2% glutamine (BioWhittaker, Walkersville, MD); RD and COS-7 cells, D-MEM, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD); and CHO cells, Ham's F-12, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The cells are incubated for either 48 or 60 hours. Cell lysates are collected as described below in Example 3. Supernatants are harvested and filtered through 0.45 µm syringe filters. Supernatants are evaluated using the using 96-well plates coated with a murine monoclonal antibody directed against HIV antigen, for example a Coulter p24-assay (Coulter Corporation, Hialeah, FL, US). The HIV-1 antigen binds to the coated wells. Biotinylated antibodies against HIV recognize the bound antigen. Conjugated streptavidin-horseradish peroxidase reacts with the biotin. Color develops from the reaction of peroxidase with TMB substrate. The reaction is terminated by addition of 4N H₂SO₄. The intensity of the color is directly proportional to the amount of HIV antigen in a sample.

Synthetic HIV Type C expression cassettes provides dramatic increases in production of their protein products, relative to the native (wild-type Subtype C) sequences, when expressed in a variety of cell lines.

B. Signal Peptide Leader Sequences

The ability of various leader sequences to drive expression was tested by transfecting cells with wild type or synthetic Env-encoding expression cassettes operably linked to different leader sequences and evaluating expression of Env polypeptide by ELISA or Western Blot. The amino acid and nucleotide sequence of various signal peptide leader sequences are shown in Table 4.

Table 4

Leader	Amino acid sequence	DNA sequence
WTnative (8_2_TV 1_C.ZA)	MRVMGTQKNCQQWWIWI LGFWMILMIC	ATGAGAGTGATGGGGACACAGA AGAATTGTCAACAATGGTGGATA TGGGGCATCTTAGGCTCTGGAT GCTAATGATTGT
WTmod (8_2_TV 1_C.ZA)	MRVMGTQKNCQQWWIWI LGFWMILMIC	ATGCGCGTGATGGGCACCCAGAA GAAC TGCCAGCAGTGGTGGATCT GGGGCATCCTGGGCTTCTGGATG CTGATGATCTGC
Tpa1	MDAMKRGGLCCVLLCGAVFVSPS AS	ATGGATGCAATGAAGAGAGGGC TCTGCTGTGCTGCTGCTGTGTG

		GAGCAGTCTCGTTCGCCCAGC GCCAGC
Tpa2	MDAMKRLCCVLLLCGAVFVSPS	ATGGATGCAATGAAGAGAGGGC TCTGCTGTGTGCTGCTGCTGTG GAGCAGTCTCGTTCGCCCAGC

35 293 cells were transiently transfected using standard methods with native and sequence-modified constructs encoding the gp120 and gp140 forms of the 8_2_TV1_C.ZA (TV1c8.2) envelope. Env protein was measured in cell lysates and supernatants using an in-house Env capture ELISA. Results are shown in Table 5 below and indicate that the wild-type signal peptide leader sequence of the TV1c8.2 can be used to efficiently express the
40 encoded envelope protein to levels that are better or comparable to those observed using the heterologous tpa leader sequences. Furthermore, the TV1c8.2 leader works in its native or sequence-modified forms and can be used with native or sequence-modified env genes. All constructs were tested after cloning of the gene cassettes into the EcoR1 and Xho1 sites of the pCMVlink expression vector.

45

Table 5

TV1c8.2 construct	Supernatant (ng)	Lysate (ng)	Total (ng)
gp140nat.wtL	532	149	681
gp140nat.tpa1	250	20	270
gp140nat.tpa2	192	34	226
gp120mod.wtLmod	6186	4576	10762
gp120mod.tpa1	6932	3808	10740
gp120mod.wtLnat	6680	4174	10854
gp140mod.wtLmod	1844	8507	10351
gp140mod.tpa1	1854	2925	4779
gp140mod.wtLnat	1532	3015	4547

The sequence-modified TV1c8.2 envelope variant gene cassettes were subcloned into
 15 a Chiron pCMV expression vector for the derivation of stable mammalian cell lines. Stable
 CHO cell lines expressing the TV1c8.2 envelope proteins were derived using standard
 methods of transfection, methotrexate amplification, and screening. These cell lines were
 found to secrete levels of envelope protein that were comparable to those observed for
 proteins expressed using the tpa leader sequences. Representative results are shown in Table
 20 6 for two cell line clone expressing the TV1c8.2 gp120; they are compared to two reference
 clones expressing SF162 subtype B gp120 derived in a similar fashion but using the tpa
 leader. Protein concentrations were determined following densitometry of scanned gels of
 semi-purified proteins. Standard curves were generated using a highly purified and well-
 characterized preparation of SF2 gp120 protein and the concentrations of the test proteins
 25 were determined.

Table 6

CHO cell line	Clone #	Expression (ng/ml)
gp120 SF162	Clone 65	921
	Clone 71	972
gp120TV1.C8.2	Clone 159	1977
	Clone 210	1920

The results were also confirmed by Western Blot Analysis, essentially as described in
 Example 3.

Example 3Western Blot Analysis of ExpressionA. HIV Type C Coding Sequences

Human 293 cells are transfected as described in Example 2 with pCMV-based vectors containing native or synthetic HIV Type C expression cassettes. Cells are cultivated for 60 hours post-transfection. Supernatants are prepared as described. Cell lysates are prepared as follows. The cells are washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO) in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) are loaded with 20 µl of supernatant or 12.5 µl of cell lysate. A protein standard is also loaded (5 µl, broad size range standard; BioRad Laboratories, Hercules, CA).

Electrophoresis is carried out and the proteins are transferred using a BioRad Transfer Chamber (BioRad Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer (Millipore), where the transfer is performed at 100 volts for 90 minutes. The membranes are exposed to HIV-1-positive human patient serum and immunostained using o-phenylenediamine dihydrochloride (OPD; Sigma).

Immunoblotting analysis shows that cells containing the synthetic expression cassette produce the expected protein at higher per-cell concentrations than cells containing the native expression cassette. The proteins are seen in both cell lysates and supernatants. The levels of production are significantly higher in cell supernatants for cells transfected with the synthetic expression cassettes of the present invention.

In addition, supernatants from the transfected 293 cells are fractionated on sucrose gradients. Aliquots of the supernatant are transferred to Polyclear™ ultra-centrifuge tubes (Beckman Instruments, Columbia, MD), under-laid with a solution of 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 28,000 rpm in a Beckman SW28 rotor. The resulting pellet is suspended in PBS and layered onto a 20-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor.

The gradient is then fractionated into approximately 10 x 1 ml aliquots (starting at the top, 20%-end, of the gradient). Samples are taken from fractions 1-9 and are electrophoresed on 8-16% SDS polyacrylamide gels. The supernatants from 293/synthetic cells give much stronger bands than supernatants from 293/native cells.

Example 4In Vivo Immunogenicity of Synthetic HIV Type C Expression CassettesA. Immunization

To evaluate the possibly improved immunogenicity of the synthetic HIV Type C expression cassettes, a mouse study is performed. The plasmid DNA, pCMVKM2 carrying the synthetic Gag expression cassette, is diluted to the following final concentrations in a total injection volume of 100 μ l: 20 μ g, 2 μ g, 0.2 μ g, 0.02 and 0.002 μ g. To overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample is brought up to 20 μ g using the vector (pCMVKM2) alone. As a control, plasmid DNA of the native Gag expression cassette is handled in the same manner. Twelve groups of four to ten Balb/c mice (Charles River, Boston, MA) are intramuscularly immunized (50 μ l per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 1.

Table 1

Group	Gag or Env Expression Cassette	Concentration of Gag or Env plasmid DNA (µg)	Immunized at time (weeks):
1	Synthetic	20	0 ¹ , 4
2	Synthetic	.2	0, 4
5	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	0.002	0, 4
6	Synthetic	20	0
7	Synthetic	2	0
10	Synthetic	0.2	0
8	Synthetic	0.02	0
9	Synthetic	0.002	0
11	Native	20	0, 4
12	Native	2	0, 4
15	Native	0.2	0, 4
13	Native	0.02	0, 4
14	Native	0.002	0, 4
16	Native	20	0
17	Native	2	0
20	Native	0.2	0
18	Native	0.02	0
19	Native	0.002	0
20	Native	0.002	0

1 = initial immunization at "week 0"

Groups 1-5 and 11-15 are bled at week 0 (before immunization), week 4, week 6, week 8, and week 12. Groups 6-20 and 16-20 are bled at week 0 (before immunization) and at week 4.

B. Humoral Immune Response

The humoral immune response is checked with an anti-HIV antibody ELISAs (enzyme-linked immunosorbent assays) of the mice sera 0 and 4 weeks post immunization (groups 5-12) and, in addition, 6 and 8 weeks post immunization, respectively, 2 and 4 weeks post second immunization (groups 1-4).

The antibody titers of the sera are determined by using the appropriate anti-HIV polypeptide (e.g., anti-Pol, anti-Gag, anti-Env, anti-Vif, anti-Vpu, etc.) antibody ELISA. Briefly, sera from immunized mice are screened for antibodies directed against the HIV proteins (e.g., p55 Gag protein, an Env protein, e.g., gp160 or gp120 or a Pol protein, e.g., p6, 5 prot or RT, etc). ELISA microtiter plates are coated with 0.2 µg of HIV protein per well overnight and washed four times; subsequently, blocking is done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100 µl of diluted mouse serum is added. Sera are tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates are washed four times and incubated with a secondary, peroxidase-coupled anti-mouse 10 IgG antibody (Pierce, Rockford, IL). ELISA plates are washed and 100 µl of 3, 3', 5, 5'- tetramethyl benzidine (TMB; Pierce) is added per well. The optical density of each well is measured after 15 minutes. The titers reported are the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.).

Synthetic expression cassettes will provide a clear improvement of immunogenicity 15 relative to the native expression cassettes.

C. Cellular Immune Response

The frequency of specific cytotoxic T-lymphocytes (CTL) is evaluated by a standard chromium release assay of peptide pulsed mouse (Balb/c, CB6F1 and/or C3H) CD4 cells. 20 HIV polypeptide (e.g., Pol, Gag or Env) expressing vaccinia virus infected CD-8 cells are used as a positive control. Briefly, spleen cells (Effector cells, E) are obtained from the mice immunized as described above are cultured, restimulated, and assayed for CTL activity against Gag peptide-pulsed target cells as described (Doe, B., and Walker, C.M., *AIDS* 10(7):793-794, 1996). Cytotoxic activity is measured in a standard ^{51}Cr release assay. Target 25 (T) cells are cultured with effector (E) cells at various E:T ratios for 4 hours and the average cpm from duplicate wells are used to calculate percent specific ^{51}Cr release.

Cytotoxic T-cell (CTL) activity is measured in splenocytes recovered from the mice immunized with HIV Gag or Env DNA. Effector cells from the Gag or Env DNA-immunized animals exhibit specific lysis of HIV polypeptide-pulsed SV-BALB (MHC 30 matched) targets cells, indicative of a CTL response. Target cells that are peptide-pulsed and derived from an MHC-unmatched mouse strain (MC57) are not lysed.

Thus, synthetic expression cassettes exhibit increased potency for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

Example 5

5 DNA-immunization of Non-Human Primates Using a
Synthetic HIV Type C Expression Cassette

Non-human primates are immunized multiple times (e.g., weeks 0, 4, 8 and 24) intradermally, mucosally or bilaterally, intramuscular, into the quadriceps using various doses (e.g., 1-5 mg) and various combinations of synthetic HIV Type C plasmids. The 10 animals are bled two weeks after each immunization and ELISA is performed with isolated plasma. The ELISA is performed essentially as described in Example 4 except the second antibody-conjugate is an anti-human IgG, g-chain specific, peroxidase conjugate (Sigma Chemical Co., St. Louis, MD 63178) used at a dilution of 1:500. Fifty µg/ml yeast extract is added to the dilutions of plasma samples and antibody conjugate to reduce non-specific 15 background due to preexisting yeast antibodies in the non-human primates.

Further, lymphoproliferative responses to antigen can also be evaluated post-immunization, indicative of induction of T-helper cell functions.

Synthetic plasmid DNA are expected to be immunogenic in non-human primates.

20 Example 6

In vitro expression of recombinant Sindbis RNA and DNA
containing the synthetic HIV Type C expression cassette

To evaluate the expression efficiency of the synthetic Pol, Env and Gag expression cassette in Alphavirus vectors, the selected synthetic expression cassette is 25 subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors. Specifically, a cDNA vector construct for *in vitro* transcription of Sindbis virus RNA vector replicons (pRSIN-luc; Dubensky, et al., *J Virol.* 70:508-519, 1996) is modified to contain a *Pme*I site for plasmid linearization and a polylinker for insertion of heterologous genes. A polylinker is generated using two oligonucleotides that contain the sites *Xba*I, *Pml*I, 30 *Apal*, *Nar*I, *Xba*I, and *Not*I (XPANXNF, and XPANXNR).

The plasmid pRSIN-luc (Dubensky et al., *supra*) is digested with *Xba*I and *Not*I to remove the luciferase gene insert, blunt-ended using Klenow and dNTPs, and purified from

an agarose gel using GeneCleanII (Biol01, Vista, CA). The oligonucleotides are annealed to each other and ligated into the plasmid. The resulting construct is digested with *NotI* and *SacI* to remove the minimal Sindbis 3'-end sequence and A₄₀ tract, and ligated with an approximately 0.4 kbp fragment from PKSSIN1-BV (WO 97/38087). This 0.4 kbp fragment 5 is obtained by digestion of pKSSIN1-BV with *NotI* and *SacI*, and purification after size fractionation from an agarose gel. The fragment contains the complete Sindbis virus 3'-end, an A₄₀ tract and a *PmeI* site for linearization. This new vector construct is designated SINBVE.

The synthetic HIV coding sequences are obtained from the parental plasmid by 10 digestion with *EcoRI*, blunt-ending with Klenow and dNTPs, purification with GeneCleanII, digestion with *SaII*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic HIV polypeptide-coding fragment is ligated into the SINBVE vector that is digested with *XhoI* and *PmtI*. The resulting vector is purified using 15 GeneCleanII and is designated SINBVGag. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVGag and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line.

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) is digested with *SaII* and *XbaI*, to remove the beta-galactosidase gene 20 insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Gag or Env gene is inserted into the pDCMVSIN-beta-gal by digestion of SINBVGag with *SaII* and *XhoI*, purification using GeneCleanII of the Gag-containing fragment after agarose gel size fractionation, and ligation. The resulting construct is designated pDSIN-Gag, and may be used directly for *in vivo* administration or formulated using any of the methods described 25 herein.

BHK and 293 cells are transfected with recombinant Sindbis RNA and DNA, respectively. The supernatants and cell lysates are tested with the Coulter capture ELISA (Example 2).

BHK cells are transfected by electroporation with recombinant Sindbis RNA.

30 293 cells are transfected using LT-1 (Example 2) with recombinant Sindbis DNA. Synthetic Gag- and/or Env-containing plasmids are used as positive controls. Supernatants and lysates are collected 48h post transfection.

Type C HIV proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic expression cassettes.

Example 7

5 In Vivo Immunogenicity of recombinant Sindbis Replicon Vectors
 containing synthetic Pol, Gag and/or Env Expression Cassettes

A. Immunization

To evaluate the immunogenicity of recombinant synthetic HIV Type C expression cassettes in Sindbis replicons, a mouse study is performed. The Sindbis virus DNA vector 10 carrying synthetic expression cassettes (Example 6), is diluted to the following final concentrations in a total injection volume of 100 μ l: 20 μ g, 2 μ g, 0.2 μ g, 0.02 and 0.002 μ g. To overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample is brought up to 20 μ g using the Sindbis replicon vector DNA alone. Twelve groups of four to ten Balb/c mice (Charles River, Boston, MA) are 15 intramuscularly immunized (50 μ l per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 2. Alternatively, Sindbis viral particles are prepared at the following doses: 10^3 pfu, 10^5 pfu and 10^7 pfu in 100 μ l, as shown in Table 3. Sindbis HIV polypeptide particle preparations are administered to mice using intramuscular and subcutaneous routes (50 μ l per site).

Table 2

Group	Gag or Env Expression Cassette	Concentration of Gag or Env DNA (µg)	Immunized at time (weeks):
1	Synthetic	20	0 ¹ , 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	0.002	0, 4
6	Synthetic	20	0
7	Synthetic	2	0
8	Synthetic	0.2	0
9	Synthetic	0.02	0
10	Synthetic	0.002	0

1 = initial immunization at "week 0"

Table 3

Group	Gag or Env sequence	Concentration of viral particle (pfu)	Immunized at time (weeks):
1	Synthetic	10 ³	0 ¹ , 4
2	Synthetic	10 ⁵	0, 4
3	Synthetic	10 ⁷	0, 4
20	Synthetic	10 ³	0
8	Synthetic	10 ⁵	0
9	Synthetic	10 ⁷	0
10	Synthetic	10 ³	0

1 = initial immunization at "week 0"

25 Groups are bled and assessment of both humoral and cellular (e.g., frequency of specific CTLs) is performed, essentially as described in Example 4.

Example 8

Identification and Sequencing of a Novel HIV Type C Variants

A full-length clone, called 8_5_TV1_C.ZA, encoding an HIV Type C was isolated and sequenced. Briefly, genomic DNA from HIV-1 subtype C infected South African patients was isolated from PBMC (peripheral blood mononuclear cells) by alkaline lysis and anion-exchange columns (Qiagen). To get the genome of full-length clones two halves were 5 amplified, that could later be joined together in frame within the Pol region using an unique Sal 1 site in both fragments. For the amplification, 200-800 ng of genomic DNA were added to the buffer and enzyme mix of the Expand Long Template PCR System after the protocol of the manufacturer (Boehringer Mannheim). The primer were designed after alignments of known full length sequences. For the 5'half a primer mix of 2 forward primers containing 10 either thymidine (S1FCSacTA 5'-GTTTCTTGAGCTCTGGAAAGGGTTAATTAC TCCAAGAA-3', SEQ ID NO:38) or cytosine on position 20 (S1FTSacTA 5' - GTTTCTTGAGCTCTGGAAAGGGTTAATTACTCTAAGAA, SEQ ID NO:39) plus Sal 1 site, were used. The reverse primer were also a mix of two primers with either thymidine or cytosine on position 13 (S145RTSalTA 5'- 15 GTTTCTTGTGCGACTTGTCCATGTATGGCTTCCCC T-3', SEQ ID NO:40 and S145RCSalTA 5'-GTTTCTTGTGCGACTTGTCCATGCATGGCTTCCCT-3' SEQ ID NO:41) and contained a Sal 1 site. The forward primer for the 3'half was also a mixture of two primers (S245FASalTA 5'-GTTTCTTGTGCGACTGTAGTCCAGGaaATATGGCAAT TAG-3' SEQ ID NO:42 and S245FGSalTA 5'- 20 GTTTCTTGTGCGACTGTAGTCCAGGgATATG GCAA TTAG-3' SEQ ID NO:43) with Sal 1 site and adenine or guanine on position 12. The reverse primer had a Not 1 site (S2_FullNotTA 5'-GTTTCTTGCAGGCCGCTGCTAGA GATTTCACACTACCA-3' SEQ ID NO:44). After amplification the PCR products were purified using a 1% agarose gel and cloned into the pCR-XL-TOPO vector via TA cloning (Invitrogen). Colonies were checked 25 by restriction analysis and sequence verified. For the full length sequence the sequences of the 5'- and 3'half were combined. The sequence is shown in SEQ ID NO:33. Furthermore, important domains are shown in Table A.

Another clone, designated 12-5_1_TV2_C.ZA was also sequenced and is shown in SEQ ID NO:45. The domains can be readily determined in view of the teachings of the specification, for example by aligning the sequence to those shown in Table A to find the corresponding regions in clone 12-5_1_TV2_C.ZA.

As described above (Example 1, Table C), synthetic expression cassettes were generated using one or more polynucleotide sequences obtained from 8_5_TV1_C.ZA or 12-5_1_TV2_C.ZA.

5 The polynucleotides described herein have all been deposited at Chiron Corporation, Emeryville, CA.

Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

Claims

1. An expression cassette comprising

a polynucleotide sequence encoding a polypeptide including an HIV *Pol* polypeptide,
5 wherein the polynucleotide sequence encoding said *Pol* polypeptide comprises a sequence
having at least 90% sequence identity to the sequence presented of Figure 8 (SEQ ID NO:30);
Figure 9 (SEQ ID NO:31) or Figure 10 (SEQ ID NO:32).

2. An expression cassette comprising

10 a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:46,
(ii) X equals Y, and (iii) Y is at least 97.

3. The expression cassette of claim 2, comprising

15 a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:47,
(ii) X equals Y, and (iii) Y is at least 144.

4. The expression cassette of claim 3, comprising

20 a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:49
or SEQ ID NO:97, (ii) X equals Y, and (iii) Y is at least 300.

5. The expression cassette of claim 4, comprising

25 a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:49,
(ii) X equals Y, and (iii) Y is 2610.

6. The expression cassette of claim 4, comprising

30 a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:97,
(ii) X equals Y, and (iii) Y is 2565.

7. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:51
5 (ii) X equals Y, and (iii) Y is 1494.

8. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:99,
10 (ii) X equals Y, and (iii) Y is 1491.

9. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:55;
15 SEQ ID NO:57; SEQ ID NO:101; SEQ ID NO:96; SEQ ID NO:134 or SEQ ID NO:135; (ii)
X equals Y, and (iii) Y is at least 60.

10. The expression cassette of claim 9, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
20 nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:55;
SEQ ID NO:57; SEQ ID NO:101; SEQ ID NO:96; SEQ ID NO:134 or SEQ ID NO:135, (ii)
X equals Y, and (iii) Y is 624.

11. An expression cassette comprising
25 a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:58;
(ii) X equals Y, and (iii) Y is 354.

12. An expression cassette comprising
30 a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:60;
(ii) X equals Y, and (iii) Y is 876.

13. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:62; (ii) X equals Y, and (iii) Y is 3015.

5

14. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:103; (ii) X equals Y, and (iii) Y is 3009.

10

15. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:64 or SEQ ID NO:66; (ii) X equals Y, and (iii) Y is 297.

15

16. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:68, (ii) X equals Y, and (iii) Y is 1965.

20

17. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:70; (ii) X equals Y, and (iii) Y is 1977.

25

18. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:72 or SEQ ID NO:105, (ii) X equals Y, and (iii) Y is at least 30.

30

19. The expression cassette of claim 18, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:72 or SEQ ID NO:105; (ii) X equals Y, and (iii) Y is 75.

5 20. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:74 or SEQ ID NO:107, (ii) X equals Y, and (iii) Y is at least 30.

10 21. The expression cassette of claim 20, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:74 or SEQ ID NO:107; (ii) X equals Y, and (iii) Y is 246.

15 22. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:76; (ii) X equals Y, and (iii) Y is 1680.

20 23. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:78; (ii) X equals Y, and (iii) Y is 1668.

25 24. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:80, SEQ ID NO:81 or SEQ ID NO:109; (ii) X equals Y, and (iii) Y is 216.

30 25. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:83; (ii) X equals Y, and (iii) Y is 93.

5 26. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:111; (ii) X equals Y, and (iii) Y is 90.

10 27. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:85, or SEQ ID NO:113; (ii) X equals Y, and (iii) Y is 579.

15 28. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:87; (ii) X equals Y, and (iii) Y is 288.

20 29. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:115; (ii) X equals Y, and (iii) Y is 287.

25 30. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:89 or SEQ ID NO:117; (ii) X equals Y, and (iii) Y is at least 30.

30 31. The expression cassette of claim 30 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:89; (ii) X equals Y, and (iii) Y is 267.

5 32. The expression cassette of claim 30 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:117; (ii) X equals Y, and (iii) Y is 261.

10 33. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:91; (ii) X equals Y, and (iii) Y is at least 30.

15 34. The expression cassette of claim 33 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:91; (ii) X equals Y, and (iii) Y is 321.

20 35. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:93 or SEQ ID NO:94; (ii) X equals Y, and (iii) Y is 309.

25 36. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:96; (ii) X equals Y, and (iii) Y is at least 60.

30 37. The expression cassette of claim 36 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:96; (ii) X equals Y, and (iii) Y is 624.

5 38. An expression cassette comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:119, SEQ ID NO:120; SEQ ID NO:121; SEQ ID NO:122; SEQ ID NO:123; SEQ ID NO:124; SEQ ID NO:125; SEQ ID NO:126; SEQ ID NO:127; SEQ ID NO:131; SEQ ID NO:132 or SEQ ID NO:133, (ii) X equals Y, and (iii) Y is at least 60.
10

15 39. The expression cassette of claim 38, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:119, SEQ ID NO:120; SEQ ID NO:121; SEQ ID NO:122; SEQ ID NO:123; SEQ ID NO:124; SEQ ID NO:125; SEQ ID NO:126; SEQ ID NO:127; SEQ ID NO:131; SEQ ID NO:132 or SEQ ID NO:133,(ii) X equals Y, and (iii) Y is at least 300.

20 40. The expression cassette of claim 39, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:123 or SEQ ID NO:124, (ii) X equals Y, and (iii) Y is 2433.

25 41. The expression cassette of claim 39, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:122, (ii) X equals Y, and (iii) Y is 2301.

30 42. The expression cassette of claim 39, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:125; (ii) X equals Y, and (iii) Y is 2517.

43. The expression cassette of claim 39, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID
NO:126 or SEQ ID NO:127, (ii) X equals Y, and (iii) Y is 2520.

5

44. The expression cassette of claim 39, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID
NO:119, (ii) X equals Y, and (iii) Y is 1377.

10

45. The expression cassette of claim 39, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID
NO:120 or SEQ ID NO:121, (ii) X equals Y, and (iii) Y is 1839.

15

46. The expression cassette of claim 39, comprising
a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous
nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID
NO:132 or SEQ ID NO:133, (ii) X equals Y, and (iii) Y is 1890.

20

47. A polynucleotide comprising the sequence depicted in SEQ ID NO:33 or
fragments derived therefrom.

48. The polynucleotide of claim 47, wherein said fragments comprise coding
sequence for the gene products selected from the group consisting of Gag, Pol, Vif, Vpr, Tat,
Rev, Vpu, Env and Nef.

49. The polynucleotide of claim 48, wherein the fragment comprises a Gag gene
product.

30

50. The polynucleotide of claim 48, wherein the fragment comprises an Env gene
product.

51. The polynucleotide of claim 50, wherein the Env gene product is gp160, gp140 or gp120.

52. A polynucleotide comprising the sequence depicted in SEQ ID NO:45 or
5 fragments derived therefrom.

53. The polynucleotide of claim 52, wherein said fragments comprise coding sequence for the gene products selected from the group consisting of Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env and Nef.

10

54. The polynucleotide of claim 53, wherein the fragment comprises a Gag gene product.

15 55. The polynucleotide of claim 53, wherein the fragment comprises an Env gene product.

56. The polynucleotide of claim 55, wherein the Env gene product is gp160, gp140 or gp120.

20 57. A polynucleotide comprising the sequence depicted in SEQ ID NO:128 or fragments derived therefrom.

58. The polynucleotide of claim 57, wherein the fragments comprise coding sequence for Env gene products gp160, gp140 or gp120.

25

59. The expression cassette of any of claims 1 to 46, further comprising one or more nucleic acids encoding one or more viral polypeptides or antigens.

30 60. The expression cassette of claim 59, wherein the viral polypeptide or antigen is selected from the group consisting of Gag, Env, vif, vpr, tat, rev, vpu, nef and combinations thereof.

61. The expression cassette of any of claims 1 to 46, further comprising one or more nucleic acids encoding one or more cytokines.

5 62. A recombinant expression system for use in a selected host cell, comprising, an expression cassette of any of claims 1 to 46, and wherein said polynucleotide sequence further comprises control elements capable of driving expression in the selected host cell.

10 63. The recombinant expression system of claim 62, wherein said control elements are selected from the group consisting of a transcription promoter, a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

15 64. The recombinant expression system of claim 62 wherein said transcription promoter is selected from the group consisting of CMV, CMV+intron A, SV40, RSV, HIV-Ltr; MMLV-ltr, and metallothionein.

20 65. A cell comprising an expression cassette of any of claims 1 to 46, and wherein said polynucleotide sequence further comprises control elements compatible with expression in the selected cell.

66. The cell of claim 65, wherein the cell is selected from the group consisting of a mammalian cell, an insect cell, a bacterial cell, a yeast cell, a plant, an antigen presenting cell, a primary cell, an immortalized cell, and a tumor derived cell.

25 67. The cell of claim 66, wherein the cell is selected from the group consisting of BHK, VERO, HT1080, 293, RD, COS-7, and CHO cells.

68. The cell of claim 67, wherein said cell is a CHO cell.

30 69. The cell of claim 66, wherein the cell is either *Trichoplusia ni* (Tn5) or Sf9 insect cells.

70. The cell of claim 66, wherein the antigen presenting cell is a lymphoid cell selected from the group consisting of macrophage, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof.

5 71. A composition for generating an immunological response, comprising an expression cassette of any of claims 1 to 46.

72. The composition of claim 71, further comprising one or more *Pol* polypeptides.

10 73. The composition of claim 72, further comprising an adjuvant.

74. A composition for generating an immunological response, comprising an expression cassette of claim 52.

15 75. The composition of claim 74, further comprising a *Pol* polypeptide.

76. The composition of claim 74, further comprising one or more polypeptides encoded by the nucleic acid molecules of claim 60.

20 77. The composition of claim 76, further comprising an adjuvant.

78. A method of immunization of a subject, comprising, introducing a composition of claim 71 into said subject under conditions that are compatible with expression of said expression cassette in said subject.

25 79. The method of claim 78, wherein said expression cassette is introduced using a gene delivery vector.

80. The method of claim 79, wherein the gene delivery vector is a non-viral vector.

30 81. The method of claim 79, wherein said gene delivery vector is a viral vector.

82. The method of claim 79, wherein said gene delivery vector is selected from the group consisting of an adenoviral vector, a vaccinia viral vector, an AAV vector, a retroviral vector, a lentiviral vector and an alphaviral vector.

5 83. The method of claim 82, wherein said gene delivery vector is a Sindbis-virus derived vector.

84. The method of claim 82, wherein said gene delivery vector is a cDNA vector.

10 85. The method of claim 82, wherein said gene delivery vector is a eukaryotic layered viral initiation system (ELVIS).

15 86. The method of claim 79, wherein said composition delivered using a particulate carrier.

87. The method of claim 79, wherein said composition is coated on a gold or tungsten particle and said coated particle is delivered to said subject using a gene gun.

20 88. The method of claim 79, wherein said composition is encapsulated in a liposome preparation.

89. The method of claim 79, wherein said subject is a mammal.

90. The method of claim 89, wherein said mammal is a human.

25 91. A method of generating an immune response in a subject, comprising:
providing an expression cassette of any of claims 1 to 46,
expressing said polypeptide in a suitable host cell,
isolating said polypeptide, and
30 administering said polypeptide to the subject in an amount sufficient to elicit an immune response.

92. A method of generating an immune response in a subject, comprising introducing into cells of said subject an expression cassette of any one of claims 1 to 46, under conditions that permit the expression of said polynucleotide and production of said polypeptide, thereby eliciting an immunological response to said polypeptide.

5

93. The method of claim 92, where the method further comprises co-administration of an HIV polypeptide.

10 94. The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out before introducing said expression cassette.

95. The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out concurrently with introducing said expression cassette.

15 96. The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out after introducing said expression cassette.

20 97. The expression cassette of claim 59, wherein the viral polypeptide or antigen is selected from the group consisting of polypeptides derived from hepatitis B, hepatitis C and combinations thereof.

25

Gag_AF110965_BW_mod

ATGGGCGCCCGCGCCAGCATCCTGCAGGGCAAGCTGGACGCGCTGGGAGCGCATCCGCC
TGCAGCCCGGGCAAGAAGTGCTACATGATGAAGCACCTGGTGTGGGCCAGCGCGAGCT
GGAGAAGTTGCCCCCTGAACCCCGCCCTGCTGGAGACCAGCGAGGGCTGCAAGCAGATCATC
CGCCAGCTGCACCCCCGCCCTGCAAGACCGGAGCGAGGAGCTGAAGAGCCTGTTAACACCG
TGGCCACCCCTGACTGCGTGCACGAGAAGATCGAGGTGCACCAAGGAGGCCCTGGA
CAAGATCGAGGAGGAGCAGAACAAAGTGCCAGCAGAAGATCCAGCAGGCCAGGCCGAC
AAGGGCAAGGTGAGCCAGAACTACCCCATCGTCAGAACCTGCAGGGCCAGATGGTGCACC
AGGCCATCAGCCCCCGCACCCCTGAACGCCCTGGGTGAAGGTGATCGAGGAGAAGGCCCTCAG
CCCCGAGGTGATCCCCATGTTACCGCCCTGAGCGAGGGGCCACCCCCCAGGACCTGAAC
ACGATGTTGAACACCGTGGGCGGCCACCAGGCCGCATGCAGATGCTGAAGGACACCATCA
ACGAGGAGGCCGCGAGTGGGACCGCGTGCACCCCGTGCACGCCGCCCATGCCCGG
CCAGATGCGCGAGCCCCCGGGCAGCGACATGCCGGCACCACCGACCCCTGCAGGAGCAG
ATGCCCTGGATGACCAGCAACCCCCCCCACATCCCCGTGGCGACATCTACAAGCGGTGATCA
TCCTGGGCCTGAACAAGATCGTGCAGGATGTACAGCCCGTGAGCATCCTGGACATCAAGCA
GGGCCCCAAGGAGGCCCTCCCGCAACTACGTGGACCGCTTCTTCAAGACCTGCCGCCGAG
CAGAGCACCCAGGAGGTGAAGAAGACTGGATGACCGACACCCCTGCTGGTGCAGAACGCCAAC
CCGACTGCAAGACCATCCTGCCGCTCTGGCCCCGGGCCAGCCTGGAGGAGATGAC
CGCCTGCCAGGGCGTGGCGGCCAGCCACAAGGCCCGTGTGGCCAGGGATGAGC
CAGGCCAACACCAGCGTGTGATGCAAGAAGAGCAACTTCAAGGGCCCCGGCATCGTCA
AGTGCCTCAACTGCGCAAGGAGGGCACATGCCCGCAACTGCCGCCCCCCGCAAGAA
GGGCTGCTGGAAGTGGCGCAAGGAGGGCACCAAGATGAAGGACTGCACCGAGGCCAGGCC
AACTTCTGGCAAGATCTGGCCAGCCACAAGGCCGCCGCAACTTCTGCAAGAGCC
GCCCGAGGCCACCGCCCCCGGCCAGAGAGCTTCCGCTTGCAGGGAGACCACCCCGGCCA
GAAGCAGGAGAGCAAGGACCGCGAGACCCCTGACCAGCCTGAAGAGCCTGTTGGCAACGAC
CCCCCTGAGCCAGTAA

Figure 1

Gag_AF110967_BW_mod.

ATGGGGCGCCCGCGGCCAGCATCCTGCGCGGCAGAAGCTGGACAAGTGGAGAAGATCCGCC
TGGCCTCCGGCGGCAAGAACGACTACATGCTGAAGCACCTGGTGTGGGCCAGCCGCGAGCT
GGAGGGCTTCGCCCTGAACCCCGGCCTGCTGGAGACCGCCGAGGGCTGCAAGCAGATCATG
AACGAGCTGCAGCCGCCCTGCAGACCGGACCGAGCTGCCAGCCTGTACAACACCG
TGGCCACCCGTACTGCGTGACGCCGCATCGAGGTCCCGACACCAAGGAGGCCCTGGA
CAAGATCGAGGAGGAGCAGAACAAAGTCCCAGCAGAACGCCAGCAGGCCAAGGAGGCCAC
GGCAAGGTGAGCCAGAACTACCCCATCGTCAGAACCTGCAGGGCCAGATGGTGCACCAGG
CCATCAGCCCCCGCACCCCTGAACGCCCTGGTGAAGGTGATCGAGGAGAAGGCCCTCAGCCC
CGAGGTGATCCCCATGTTCACCGCCCTGAGCGAGGGGCCACCCCGAGGACCTGAACACG
ATGTTGAACACCGTGGCGGCCACCAGGCCATGCAGATGCTGAAGGACACCATCAACG
AGGAGGCCCGAGTGGGACCGCCTGCACCCCGTGAGGCCGGCCCTGGGCCCCGGCCA
GATGCGCGACCCCCCGCGCAGCGACATGCCGGCGCCACCGCACCCCTGCAGGAGCAGATC
GCCTGGATGACCAGCAACCCCCCGTGGCGACATCTACAAGCGGTGGATCATCC
TGGGCCTGAACAAGATCGTGCAGATGTACAGCCCCGTGAGCATCCTGGACATCCGCCAGGG
CCCCAAGGAGCCCTCCGCGACTACGTGGACCGCTTCTCAAGACCTGCGCCGAGCAG
GCCACCCAGGACGTGAAGAACTGGATGACCGAGACCTGCTGGTGCAGAACGCCAACCCCG
ACTGCAAGACCATCCTGCGCGCTCTCGGCCCGGCCACCCCTGGAGGAGATGATGACCGC
CTGCCAGGGCGTGGCGGCCACAGATGAAGGACTGCACCGAGGCCAG
GCCAACAGCGTGAACATCATGATGCAGAAGAGCAACTCAAGGGGCCGGCGAACGTCA
AGTGCTTCAACTGCGGCAAGGAGGGCCACATGCCAACAACTGCCGCCCGCGAACGAA
GGGCTGCTGGAAGTGGCGAACGGAGGGCCACCGAGATGAAGGACTGCACCGAGGCCAGGCC
AACTTCTGGCAAGATCTGGCCCAGCCACAAGGGCCGGCAACTTCCCTGCAGAAC
GCAGCGAGCCCGCCGCCACCGTGGCCACCGGCCGGAGAGCTCCGCTTCGA
GGAGACCACCCCCGCCACCGAGGCCAACGGACCGAGGCCCTACCGCGAGGCCCTG
ACCGCCCTGCGCAGCCTGTTGGCAGCGGCCCTGAGCCAGTAA

Figure 2

Fig. 3**Env_AF110968_C_BW_opt**

--> signal peptide (1-81)

ATGCGCGTGTATGGGCATCCTGAAGAACCTACCAAGCAGTGGTGGATGTGGGGCATCCTGGGCTCTGGATGCTGATCA
 \/\--> gp120/140/160 (82)
 TCAGCAGCGTGGTGGCACCTGTGGGTGACCGTGTACTACGGCGTGCCGTGGAAGGAGGCCAAGACCACCC
 GTTCTGCACCAGCGACGCCAACGGCTACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCCTGCGTGCCCACC
 GACCCCCAACCCCCCAGGGAGATCGTGTGGAGAACGTGACCGAGAACCTCAACATGTGGAAGAACGACATGGTGGACC
 AGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGGCCCTGCGTGAAGCTGACCCCCCTGTGCGTGAC
 CCTGAAGTGCCGCAACGTGAACGCCACCAACAAACATCAACAGCATGATCGACAACAGCAACAAGGGCGAGATGAAG
 AACTGCAGCTCAACGTGACCACCGAGCTGCCGACCGCAAGCAGGAGGTGCACGCCCTGTTCTACCGCCTGGACG
 TGGTGCCCTGCAGGGCAACACAGCAACGAGTACCGCCTGATCAACTGCAACACCCAGGCCATACCCAGGCC
 CCCCCAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCACCCCGCCGGCTACGCCATCCTGAAGTGCAACAAAC
 CAGACCTTCAACGGCACCGGCCCTGCAACAACGTGAGCAGCGTGCAGTGCAGGCCACGGCATCAAGCCGTGGTA
 GCACCCAGCTGCTGCTGAACGGCAGCCTGGCCAAGGGCGAGATCATCATCCGAGCGAGAACCTGGCAACACGC
 CAAGATCATCATCGTGCAGCTGAACAAGCCGTGAAGATCGTGTGCGCCCCAACAAACAACACCCGCAAGAGC
 GTGCGCATCGGCCCCGGCCAGACCTTCTACGCCACCGCCAGATCATGGGACATCGGCCAGGCCACTGCA
 TCAACAAAGACCGAGTGGAACAGCACCCCTGCAGGGCGTGAGCAAGAACGCTGGAGGAGCACCTCAGCAAGAACGCCAT
 CAAGTTGAGGCCAGCAGCGGGCGACCTGGAGATCACCACCCACAGCTTCAACTGCCGCGGAGTTCTTCTAC
 TGCGACACCAGCAGCTGTTCAACAGCACCTACAGCCCCAGCTTCAACGGCACCGAGAACAGCTGAAACGGCACCA
 TCACCATCACCTGCCCATCAAGCAGATCATCAACATGTGGCAGAACGGCCGCGCCATGTACGCCCTGGGCG
 CGCCGGCAACCTGACCTGCGAGAGCAACATCACCGCCTGCTGCTGACCCGCGACGGCGCAAGACCGGCCAAC
 GACACCGAGATCTCCGCCCCGGCGCGACATGCGCACAACGAGCTGACAGCTGACAGCTGACAGCTGAGG
 TGGAGATCAAGCCCTGGCGTGGCCCCCACCGAGGCCAGCGCCGCGTGGTGGAGCGCGAGAACGGCG
 CATCGGCCGTGTTCTGGCTTCCTGGCGCCGGCAGCACCATGGCGCCGCCAGCATCACCTGACCGTGC
 CAGGCCGCGCTGCTGAGCGGCATCGTCAGCAGCAGAACACCTGCTGCGGCCATCGAGGCCAGCAGCACC
 TGCTGCGACTGACCGTGTGGGCATCAAGCAGCTGCGAGACCCGATCCTGGCGTGGAGCGCTACCTGAAGGACCA
 CGAGCTGCTGGGCATCTGGGCTGCGAGCGCAAGCTGATCTGCACCAACCGCCGTGCCCTGGAACAGCAGCTGGAGC
 AACCGCAGCCACGAGATCTGGAACACATGACCTGGATGCAGTGGAGCCGAGATCAACAACTACACCGACA
 CCATCTACCGCCTGCTGGAGGGAGGCCAGAACAGCAGGAGAACAGAGAACGGACCTGCTGGCCCTGGACAGCTG
 gp140 (2025) <--\/
 GCAGAACCTGTTGAACTGGTCAGCATCACCAACTGGCTGTTGACATCAAGATCTCATGATCGTGGCG
 CTGATCGCCCTGCCATCATCTCGCCGTGCTGAGCATCGTAACCGCGTGCAGGCCAGGGCTACAGCCCCCTGCC
 TCCAGACCCCTGACCCCAACCCCCCGCGAGCCGACCGCCCTGGGCCATCGAGGAGGGAGGGCGCGAGCAGGACCG
 CGGCCGAGCATCGCCCTGGTGAAGCGCTTCTGGCCCTGGCGTGGAGCAGCTGCGCAGCCGTGCGCTGTCAGC
 TACCAACCGCCTGCCGACTTCATCCTGATCGCCGCCCGCGTGGAGCTGCTGGGCCAGCGCCGCGTGGAGGCC
 TGAAGTACCTGGGAGCCCTGGTGCAGTACTGGGCTGGAGCTGAGAACAGAGGCCATCACGCTGCTGGACACCAT
 CGCCATCGCCGTGGCGAGGGCACCGACCGCATCATCGAGTTCATCCAGCGCATCTGCCGCGCATCCGAAACATC
 gp160, gp41 (2547) <--\/
 CCCCGCCGATCCGCCAGGGCTTCGAGGCCGCCCTGCAAGTAA

Fig. 4

Env_AF110975_C_BW_opt

--> signal peptide (1-72)
 ATGCGCGTGCAGCAGCTGGCAGCAGTGGTGATCTGGGCATCCCTGGCTCTGGATCTGCAGCG
 gp120/140/160 (72)
 GCCTGGCAACCTGTGGTGACCGTGTACGACGGCGTCCCCGTGGCGAGGCCAGCACCACCTGTTCTGC
 CAGCGACGCCAAGGCCTACGAGAAGGAGGTGCACAACGTGTGGCCACCCACGCCCTGCGTGCCCACCGACCCAA
 CCCCAGGAGATCGAGCTGGACAACGTGACCGAGAACCTCAACATGTGAAGAACGACATGGTGACCGAGATGCACG
 AGGACATCATCAGCCTGTGGGACCAAGAGCTGAAGCCCCCGTGAAGCTGACCCCCCTGTGCGTGACCCCTGAAGT
 CACCAACTACAGCACCAACTACAGCAACACCATGAACGCCACCAGCTACAACAACACACCACCGAGGAGATCAAG
 AACTGCACCTTCAACATGACCACCGAGCTGCGGACAAGAACGAGCAGCAGGTGTACGCCCTGTTCTACAAGCTGGACA
 TCGTGCCTGAACAGCAACAGCAGTAGCCCTGATCAACTGCAACACCAAGGCCATACCCAGGCC
 CAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCCCCCGCCGCTACGCCATCTGAAGTGCAAGAACAAAC
 ACCAGCAACGGCACCGGCCCCCTGCCAGAACGTGAGCACCGTGCAGTGACCCACGGCATCAAGCCGTGGTGGAGCA
 CCCCCCTGCTGCTGAACGGCAGCCTGGCGAGGGCGGGAGATCATCATCCGAGCAAGAACCTGAGCAACAAACGC
 CTACACCATCATCGTGCACCTGAACGACAGCGTGGAGATCGTGTGACCCGCCAACAAACAACACCCGCAAGGGC
 ATCCGCATCGGCCCCGGCCAGACCTCTACGCCACCGAGAACATCATCGGCACATCGGCCAGGCCACTGCAACA
 TCAGCGCCGGCGAGTGGAAACAAGGCCGTGAGCGCGTGGAGCGCAAGCTGCCGAGCAGTCCCAACAAGACCAT
 CGAGTCCAGCCAGCAGCGCGGCGACCTGGAGATCACCAACCCACAGCTTCAACTGCCGCGGAGTTCTTCTAC
 TGCAACACCAGCAAGCTGTTAACAGCAGCTAACAGGCCACAGCTACCGCGGACCGAGAGCAACAGCAGCATCA
 TCACCCCTGCCCTGCCATCAAGCAGATCATCGACATGTGGCAGAAGGTGGGCCGCATCTACGCC
 CGAGGGCAACATCACCTGCAGCAGCATCACGGCCTGCTGGCCCGCAGCAACATGGGCCGCCAGCATCACCC
 ACCGAGATCTCCGCCCCAGGGCGGAGATGAAGGACAACCTGGCGAACAGAGCTGAGTACAAGTACAAGGTGG
 gp120 (1509) <--\--> (1510) gp41
 AGATCAAGCCCCCTGGCGTGGCCCCACCGAGGCCAGCGCCGCTGGAGCGCAGAACAGCGCCGTGGCAT
 CGGGCGCGTGTCTGGCTTCTGGGCCCGCCGGCAGCAACATGGGCCGCCAGCATCACCC
 GCCCGCCAGCTGCTGAGCGCATCGTGCAGCAGCACCGAGCAACCTGCTGCCGCATCGAGGCCAGCAGCACATGC
 TGAGCTGACCGTGTGGGCATCAAGCAGCTGCGAGGCCCGTGTGGCATCGAGCGTACCTGAAGGACAGCA
 GCTGCTGGCATCTGGGCTGAGCGGCAAGCTGATCTGCACCAACCACCGTGGCCCTGGAACAGCAGCTGGAGCAAC
 AAGACCCAGGGCAGAGATCTGGGAGAACATGACCTGGATGCAGTGGACAAGGAGATCAGCAACTACACCGC
 TCTACCGCCTGCTGGAGGAGGCCAGAACAGCAGGAGCAGAACAGAGAACGGACCTGCTGGCCCTGGACAGCC
 CAACCTGTGGAGCTGGTCAACATCAGCAACTGGCTGTGGTACATCAAGATCTTCACTGATCGTGGCGGC
 ATCGGCCCTGCCATCATCTCGCCGTGCTGAGCATCGTAACCGCGTGGCCAGGGCTACAGCC
 AGACCCCTGACCCCAACCCCGCGGCCCTGGACCGCCCTGGCCGATCGAGGAGGAGGGCGGCAGCAGGACCG
 CCGCAGCATCCGCCTGGTGCAGGGCTTCTGGCCCTGGCGTGGACGACCTGCCAGCCTGTGCCTGTT
 CACCGCCTGCCGACCTGATCCTGGTACCGCCCGTGGCCGATCGAGGAGGAGGGCGGCAGCAGGCC
 AGCGCGGCTGGAGGCCCTGAAGTACCTGGCAGCCTGGTGCAGTACTGGGCCCTGGAGCTGAAGAACAGGCC
 CAGCCTGCTGGAGCAGCATGCCATGCCGTGGCCAGGGCAGCCGACCTGAGCTACAGGAGGTGATCCAGGCC
 CGCGCCTCTGCAACATCCCCCGCCGCGTGCAGGCCAGGGCTTCGAGGCCGCCTGCAGTAA

G~~g~~_AF110965_BW_opt
ATGGGCGCCCGGCCAGCATCCTGCAGCGGCTGGACGCCATCCGCCTGCGCCCCGG
CGGCAAGAACTGCTACATGATGAAGCACCTGGTGTGGCCAGCCGAGCTGGAGAAGTTGCCCCCTGAACC
CCGGCCTGCTGGAGACCAGCGAGGGCTGCAAGCAGATCATCCGCCAGCTGCACCCCGCCCTGCAGACCGGC
AGCGAGGAGCTGAAGAGCCTGTTAACACCGTGGCCACCCCTGTAUTGCCTGCACGAGAAAGATCGAGGT[G]
CGACACCAAGGAGGCCCTGGACAAGATCGAGGAGGAGCAGAACAGTGCAGCAGAACAGATCCAGCAGGCCG
AGGCCGCGACAAGGGCAAGGTGAGCCAGAACACTACCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCAC
CAGGCCATCAGCCCCCGCACCCCTGAACGCCCTGGGTGAAGGTGATCGAGGAGAACGGCCTTCAGCCCCGAGGT
GATCCCCATGTTACCGCCCTGAGCGAGGGGCCACCCCCCAGGACCTGAACACCATGCTGAACACCGTG[G]T
GCGGCCACCAAGGCCGCGCATGCAGATGCTGAAGGACACCATCACAGAGGAGGCCGAGTGGGACCGCGTG
CACCCCGTGCACGCCGGCCCCATGCCCGGCCAGATGCCGAGCCCCGGCAGCGACATGCCGGCAC
CACCAAGCACCCCTGCAGGAGCAGATGCCGTGGATGACCAGAACCCCCCATTCCCCTGGCGACATCTACA
AGCC[G]CTGGATCATCCTGGCCTGAACAAGATCGTGC[G]CATGTACAGCCCCGTGAGCATCCTGGACATCAAG
CAGGGCCCCAAGGAGCCCTTCCCGACTACGTGGACCGCTTCTCAAGACCCCTGCCGCCAGCAGAGCAC
CCAGGAGGTGAGAACTGGATGACCGACACCCCTGCTGGTGCAGAACGCCAACCCCCACTGCAAGACCATCC
TGGCGC[G]CTGGGCCCCGGGCCAGCCTGGAGGAGATGATGACCGCC[G]GCCAGGGCGTGGGCGGGCCCCAGC
CACAAAGGCCCGTGCCTGGCGAGG[G]CATGAGCCAGGCCAACACCAGCGTATGAGAACAGAGCAACTT
CAAGGGCCCC[G]GATCG[G]AAGTGCTTCACTGCCAGGAGGGCACATGCCCGCAACTGCCGOG
CCCCCGCAAGAAGGGCTGCTGGAAAGTGCAGGCCAGGAGGGCACCAAGGAGGGCACATGCCCGCAACTGCCGAG
GCCAACCTCTGGCAAGATCTGCCAGGCCAACAGGAGGGCACATGCCCGCAACTTCTGCAGAGGCCAGGGCCCCAG
GCCAACCCCCCCCCGGCGAGAGCTTCCGCTTCGAGGAGACCACCCCGGCCAGAACAGCAGGAGAGCAAGG
ACCGCGAGACCCCTGACCGCCTGAAGAGCCTGTCGGCAACGACCCCTGAGCCAGTAA

Figure 5

Gag_AF110967_BW_opt
ATGGGCGGCCGCAGCATCCTGCGGGCGAGAAGCTGGACAAGTGGAGAAGATCCGCCTGCGCCCCGG
CGGCAAGAACGCACTACATGCTGAAGCACCTGGTGTTGGGCCAGCCGAGCTGGAGGGCTTCGCCCTGAA
CCGGCCCTGCTGGAGACCGCCGAGGGCTGCAAGCAGATCATGAAGCAGCTGCAGCCGCCCTGAGACCGGC
ACCGAGGAGCTGCCAGCCGTACAAACACCGTGGCCACCCCTGTACTGCGTGCACGCCGGCATCGAGG
CGACACCAAGGAGGCCCTGGACAAGATCGAGGAGGAGCAGAACAGAGCAGCAGAAGACCCAGCAGGCCA
AGGAGGCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGCCAGATGGTGCACCA
GCCATCAGCCCCCGCACCTGAACGCCCTGGGTGAAGGTATCGAGGAGAAGGCCCTCAGCCCCGAGGTGAT
CCCCATGTTCACCGCCCTGAGCAGGGGCCACCCCCCAGGACCTGAACACCATGCTGAACACCGTGGGCT
GOCACCAGGCCCOATGCAGATGCTGAAGGACACCATCAACGAGGAGGCCGAGTGGGACCGCCTGCAC
CCCGTGCAGGCCGGCCCGTGGCCCCCGCCAGATGCGCAGCCCCCGCCAGOGACATGCCGGGCCAC
CAGCACCCCTGCAGGAGCAGATGCCCTGGATGACCAGCAACCCCCCGTGGCGACATCTACAGC
GGTGGATCATCCTGGGCTGAACAAGATCGTGCCTGATGTACAGCCCGTGAGCATCCTGGACATCCGCCAG
GGCCCGAAGGAGGCCCTCCGCACTACGTGGACCGCTTCTCAAGACCCCTGCGGCCGAGCAGGCCACCA
GGACGTGAGAACGACTGGATGACCGAGACCCCTGCTGGTGAGAACGCCAACCCGACTGCAAGACCATCCTGC
GCCGCTGGCCCCGGCCACCCCTGGAGGGAGATGATGACCGCCCTGCCAGGGCTGGCGGGCCCGGCCAC
AAGGCCCGGTGCTGGCCAGGGCATGAGCCAGGCCAACAGCGTGAACATCATGATGGAGAACGAGCAACTT
CAAGGGCCCCCGCAACGTCAGTGCTTAACGTGGCAAGGAGGGCCACATGCCAAGAACGACTGCCCG
CCCCCGCAAGAAGGGCTGCTGGAGTGCGGCAAGGAGGGCCACAGATGAAGGACTGCCACCGAGGCCAG
GCCAACCTCCTGGCAAGATCTGGCCAGCCACAAGGGCCGCCCCGGCAACTCCTGCAGAACGCCAGCGA
GCCCGCCGCCACCGTGCCACCGGCCGCCCCGGCCAGAGGCTTCCGCTTCGAGGGAGACCAACCCCGCCC
CCAAGCAGGAGGCCAAGGACCGCGAGCCCTACCGCGAGGCCCTGACCGCCCTGCCAGCCCTGTTGGCAGC
GGCCCCCTGAGCCAGTAA

Figure 6

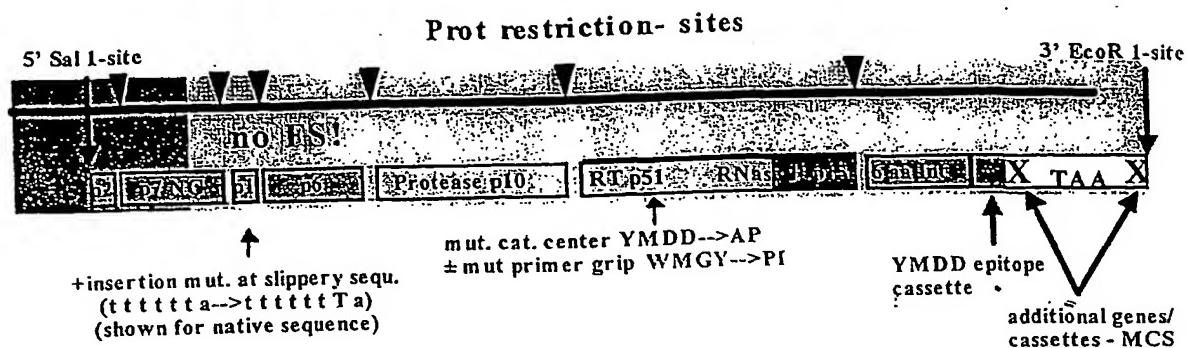


FIGURE 7

PR975(+) (SEQ ID NO:30)

GTCGACGCCACCATTGGCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGAT
GCAGCGCAGCAACTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAA
GGAGGGCCACATCGCCCGCAACTGCCGCCCGCAAGAAGGGCTGCTGGAAGT
GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTC
CGCGAGGACCTGGCTTCCCCAGGGCAAGGCCGAGTTCCCAGCGAGCAGAA
CCGCGCCAACAGCCCCACCAGCCCGAGCTGCAGGTGCGCGACAACCCCCGCA
GCGAGGCCGGCGCCGAGGCCAGGGCACCTGAACCTCCCCAGATCACCCCTGTGGC
AGCGCCCCCTGGTGAACATCAAGGTGGGGGCCAGATCAAGGAGGCCCTGCTGGAC
ACCGGCCGACGACACCGTGTGGAGGAAGATGAGCCTGCCGGCAAGTGGAAAGCC
CAAGATGATGGCGGATGGCGCTTCAAGGTGCGCCAGTACGACCAGATCCT
GATCGAGATCTGCGGCAAGAAGGCCATGGCACCCGTGTGATCGGCCCCACCCCCGT
AACATCATCGGCCGAAACATGCTGACCCAGCTGGCTGCACCCCTGAACCTCCCCAT
CAGCCCCATCGAGACCGTGGCGTAAGCTGAAGGCCGGCATGGACGGCCCCAAGG
TGAAGCAGTGGCCCCCTGACCGAGGGAGAAGATCAAGGCCGTGACCGCCATCTGCGAG
GAGATGGAGAAGGAGGGCAAGATACCAAGATCGGCCCCGAGAACCCCTACAACAC
CCCCGTGTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT
TCCCGAGCTGAACAAGCGCACCCAGGACTCTGGGAGGTGCACTGGGATCCCC
ACCCCGCCGGCTGAAGAAGAAGAGCGTGAACCGTGTGGACGTGGCGACGCC
TACTTCAGCGTGGCCATCGAGGACTTCCGCAAGTACACCGCCTTACCATCCCC
AGCATCAACAACGAGACCCCGCATCCGCTACCAAGTACAACGTGCTGCCAGGGC
TGGAAAGGGCAGCCCCAGCATCTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTC
CGCGCCCGCAACCCCGAGATCGTGTACCTACAGTACATGGACGACCTGTACGTGGC
AGCGACCTGGAGATCGGCCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCT
GCTGCGCTGGGCTTCAACCACCCCGACAAGAACGACCCAGAACGGAGCCCCCTCCT
GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCACTGGGAGCTGCG
CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAACGACCTGGGAGCTGCG
GGGCCAGCCAGATCTACCCCGCATCAAGGTGCCAGCTGTGCAAGCTGCTGCG
GCGCCAAGGCCCTGACCGACATCGTCCCCGTGACCGAGGGAGGCCAGCTGGAGCTG
GCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCAG
CAAGGACCTGGTGGCCAGATCCAGAACGACGGGACGACCAAGTGGACCTACAGA
TCTACCAAGGAGCCCTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACC
GCCACACCAACGACCGTGAAGCAGCTGACCGAGGCCGTGAGAACGATGCCATGGA
GAGCATCGTGTACCTGGGCAAGAACCCCCAAGTCCGCTGCCATCCAGAACGGAGAC
CTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCAGTGGAGTT
CGTGAACACCCCCCTGGTGAAGCTGGTACCGACTGGGAGACCAAGATCGGCA
CGCGCCGAGACCTTCACTGGACGGCGCCGCAACCGCGAGACCAAGATCGGCA
AGGCCGGCTACGTGACCGACGGGGCCGGCAGAACGAGATCGTGAACGCTGACCGAGACC
ACCAACCGAACGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGCAG
CGAGGTGAACATCGTGAACGACGCCAGTACGCCCTGGCATCCAGGCCAGGCC
CGACAAGAGCGAGAGCGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAAGG
AGAAGGTGTACCTGAGCTGGGTGCCGCCACAAGGGCATGGCGGCAACGAGCAG
ATCGACAAGCTGGTGGACCAAGGGCATCCGAAGGTGCTGTTCTGGACGGCATCGAT
GGCGGACATCGTGTACCTACCAAGTACATGGACGACCTGTACGTGGCAGCGGCGGCC
AGGATCGATAAAAGCTTCCCCGGCTAGCACCCTGTGAATT

FIGURE 8

PR975YM (SEQ ID NO:31)

GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGAT
GCAGCGCAGCAACTCAAGGGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAA
GGAGGGCCACATCGCCCGCAACTGCCGCCCGCAAGAAGGGCTGCTGGAAGT
GCGGCAAGGAGGGCCACCAAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTC
CGCGAGGACCTGGCTTCCCCAGGGCAAGGCCGAGTTCCCAGCGAGCAGAA
CCGCGCCAACAGCCCCACCAAGCCCGAGCTGAGGTGCGCGGACAACCCCCGCA
GCGAGGCCGGCGCCAGCGCCAGGGCACCCCTGAACCTCCCCAGATCACCCCTGTTG
AGCGCCCCCTGGTGAACATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC
ACCGGCACGACACCGTGTGGAGGAGATGAGCTGCCGGCAAGTGGAA
CAAGATGATCGCGGCATCGCGCTTCATCAAGGTGCGCCAGTACGACCAAGATCCT
GATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGTGATCGGCCAACCCCGT
GAACATCATCGGCCGAAACATGCTGACCCAGCTGGCTGCACCCCTGAACCTCCCCAT
CAGCCCCATCGAGACCGTGTGGCAAGCTGAAGGCCGGATGGACGGCCCCAAGG
TGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG
GAGATGGAGAAGGAGGGCAAGATACCAAGATCGGCCCGAGAACCCCTACAACAC
CCCCGTGTTGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT
TCCCGAGCTGAACAAGCGACCCAGGACTCTGGAGGTGCAAGCTGGGATCCCC
ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGTGGACCGTGTGGACGTGGCGACGCC
TACTTCAGCGTGTGGACGAGGACTTCCGCAAGTACACCGCCTTACCATCCCC
AGCATCAACAACGAGACCCCCGGCATCCGCTACCAAGTACAACGTGCTGCCCAAGG
TGGAAAGGGAGCCCGCATCTCCAGAGCAGCATGACCAAGATCTGGAGGCCCTTC
CGCGCCCGCAACCCCGAGATCGTGTACCGTACAGGCCCTGTACGTGGGAGCGAC
CTGGAGATCGGCCAGCACCGGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCG
CTGGGGCTTACCAACCCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGTGGAT
GGGCTACGAGCTGCACCCGACAAGTGGACCGTGAGCCCATCGAGCTGCCAGA
AGGAGAGCTGGACCGTAACGACATCCAGAAGCTGGGGCAAGCTGAACCTGGG
AGCCAGATCTACCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGGCC
AAGGCCCTGACCGACATCGTGTGGACCGTGACGGAGGAGCTGGAGCTGGCG
GAACCGCGAGATCCTGCGCGAGCCCGTGACGGCGTGTACTACGACCCAGCAAGG
ACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAAGTGGACCTACAGATCTAC
CAGGAGCCCTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCA
CACCAACGACGTGAAGCAGCTGACCGAGGCCGTGAGAAGATGCCATGGAGAGCA
TCGTGATCTGGGCAAGACCCCAAGTTCGCTGCCATCCAGAAGGAGACCTGG
AGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGITCGTGA
ACACCCCCCTGGTGAAGCTGTGGTACCAAGCTGGAGAAGGAGCCATCATGGCG
CCGAGACCTTACGTGGACGGCGCCAAACCGCGAGACCAAGATCGCAAGGCC
GGCTACGTGACCGACCCGGCCCGAGAAGATCGTGTGGCTGACCGAGACCA
CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAAGGACAGCGGAGCGAGG
TGAACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCGACA
AGAGCGAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAAG
GTGTACCTGAGCTGGGTGCCGCCACAAGGGCATCGGCGGCAACGAGCAGATCGA
CAAGCTGGTGAAGGGCATCCGCAAGGTGCTGTTCTGGACGGCATCGATGGCG
GCATCGTGAATCACCAGTACATGGACGACCTGTACGTGGGAGCGGGCGGCCCTAGGA
TCGATTAAAAGCTTCCGGGCTAGCACCAGTGAATT

FIGURE 9

PR975YMWM (SEQ ID NO:32)

GTGACGCCACCATGGCGAGGCCATGAGCCAGGCCACCAGGCCAACATCCTGAT
GCAGCGCAGCAACTCAAGGGCCCCAAGCGCATCATCAAGTGCTCAACTCGGGCAA
GGAGGGCACATCGCCCGCAACTGCCGCCCGCAAGAAGGGCTGCTGGAAGT
CGGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACCTTC
CGCGAGGACCTGGCTTCCCCAGGGCAAGGCCCGAGTCCCAGCGAGCAGAA
CCGCGCCAACAGCCCCACCAGCCCGAGCTGCAGGTGCGCGGACAACCCCGCA
GCGAGGCCGGCGAGGCCAGGGCACCTGAACCTCCCCAGATCACCCCTGTGGC
AGCGCCCCCTGGTAGCATCAAGGTGGCGGCCAGATCAAGGAGGCCCTGCTGGAC
ACCGGCCCGACGACACCGTGTGGAGGAGATGAGCTGCCGGCAAGTGGAAAGCC
CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCT
GATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGTGATCGGCCCCACCCCGT
GAACATCATCGGCCGCAACATGCTGACCCAGCTGGCTGACCCCTGAACCTCCAT
CAGCCCCATCGAGACCGTGGCGGTGAAGCTGAAGGCCAGTGGACGCCATCTGCGAG
TGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG
GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC
CCCCGTGTCGCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGACT
TCCCGCAGCTGAACAAGCGCACCCAGGACTCTGGGAGGTGAGCTGGCATCCCC
ACCCCGCCGGCCTGAAGAAGAAGAGCGTGACCGTGTGGACGTGGCGACGCC
TACCTCAGCGTGCCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC
AGCATCAACAACGAGACCCCGGATCCGTAACAGTACAACGTGCTGCCAGGGC
TGGAGGGCAGCCCCAGCATCTCCAGAGCAGCATGACCAAGATCCTGGAGGCCCTTC
CGCGCCCGCAACCCCGAGATCGTGTACCGAGGCCCTGTACGTGGCAGCGAC
CTGGAGATCGGCCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCG
CTGGGGCTTCACCAACCCCGACAAGTGGACCGTGCAGCCATCGAGCTGCCAGAGAAGGAGA
CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCATCGAGCTGCCAGAGAAGGAGA
GCTGGACCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGCCAGGCCAG
ATCTACCCCGGATCAAGGTGCGCAGCTGTGCAAGCTGCTGCCGGGCCAGGCC
CTGACCGACATCGTGGCCCTGACCGAGGAGGCCAGCTGGACCTACCAAGATCTACCAAGGAGC
CGAGATCTCGCGAGCCCGTGCACGGCGTGTACTACGACCCAGCAAGGACCTGGT
GGCCGAGATCCAGAAGCAGGGCACGACCAAGTGGACCTACCAAGATCTACCAAGGAGC
CTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAAC
GACGTGAAGCAGCTGACCGAGGCCGTGCAAGAAGATGCCATGGAGAGCATCGTGT
CTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAAGGAGACCTGGAGACCT
GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGAGTTCGTGAACACCC
CCCCCTGGTGAAGCTGGTACCAAGCTGGAGAAGGAGGCCATCATGGCGCGAG
ACCTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGCAAGGCCGGCTA
CGTACCGACCGGGGCCAGAAGATCGTGTACCGTGCAGGAGACCGGGCAGCGAGGTGAAC
AGACCGAGCTGAGGCCATCCAGCTGGCCCTGCAAGGAGGCCAGGCCACCAAGAG
ATCGTACCGACGCCAGTACGCCCTGGCATCATCCAGGCCAGGCCACCAAGAG
CGAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGT
ACCTGAGCTGGTGGCCACAGGGCATCGCGGCCAGGAGCAGATCGACAAG
CTGGTAGCAAGGGCATCGCAAGGTGCTTCCCTGGACGGCATCGATGGCGGCATC
GTGATCTACCAAGTACATGGACGACCTGTACGTGGCAGCGGGGCCCTAGGATCGAT
TAAAAGCTCCGGGCTAGCACCGGTGAATT

FIGURE 10

8_5_ZA (SEQ ID NO:33)

1 TGGAAAGGGTT AATTTACTCC AAGAAAAGGC AAGAAATCCT TGATTGTGG GTCTATCACA
 61 CACAAGGCTT CTTCCCTGAT TGGCAAAACT ACACACCAGG GCCAGGGGTC AGATATCCAC
 121 TGACCTTG ATGGTGTAC AAGCTAGTGC CAGTTGACCC AGGGGAGGTG GAAGAGGCCA
 181 ACGGAGGAGA AGACAACGTG TTGCTACACC CTATGAGCCA ACATGGAGCA GAGGATGAAG
 241 ATAGAGAAGT ATTAAAGTGG AAGTTTGACA GCCTCCTAGC ACCAGACAC ATGGCCCGCG
 301 AGCTACATCC GGAGTATTAC AAAGACTGCT GACACAGAAC GGACTTTCCG CCTGGGACTT
 361 TCCACTGGGG CGTTCCGGGA GGTGTGGTCT GGGCGGGACT TGGGAGTGGT CAACCCTCAG
 421 ATGCTGCATA TAAGCAGCTG CTTTCGCCT GTACTGGGTC TCTCTCGGTG GACCAGATCT
 481 GAGCCTGGGA GCCCTCTGGC TATCTAGGGA ACCCACTGCT TAAGCCTCAA TAAAGCTTGC
 541 CTTGAGTGTCT TTAAGTAGTGTG TGTTGCCCCATC TGTTGTGTGA CTCTGGTAAC TAGAGATCCC
 601 TCAGACCCCTT TGTGGTAGTG TGAAAATCT CTAGCAGTGG CGCCCGAACCA GGGACCAGAA
 661 AGTGAAGATG AGACCAAGAGG AGATCTCTCG ACGCAGGACT CGGCTTGCTG AAGTGCACAC
 721 GGCAAGAGGC GAGAGGGGCG GCTGGTGAGT ACGCCAATT TACTTGACTA GCGGAGGCTA
 781 GAAGGAGAGA GATGGGTGCG AGAGCGTCAA TATTAAGCGG CGAAAAATTAA GATAAATGGG
 841 AAAGAATTAG GTTAAGGCCA GGGGGAAAGA AACATTATAT GTTAAAACAT CTAGTATGGG
 901 CAAGCAGGGG GCTGGAAAGA TTGCACTTA ACCCTGGCCT GTTAGAAACAA TCAGAAGGCT
 961 GTAAACAAAT AATAAAACAG CTACAACCCAG CTCTTCAGAC AGGAACAGAG GAACTTAGAT
 1021 CATTATTCAA CACAGTAGCA ACTCTCTATT GTGTACATAA AGGGATAGAG GTACGAGACA
 1081 CCAAGGAAGC CTTAGACAAG ATAGAGGAAG AACAAAACAA ATGTCAGCAA AAAGCACAAC
 1141 AGGAAAAGC AGCTGACGAA AAGGTCAGTC AAAATTATCC TATAGTACAG AATGCCAAG
 1201 GGCAAATGGT ACACCAAGCT ATATCACCTA GAACATTGAA TGCATGGATA AAAGTAATAG
 1261 AGGAAAAGGC TTTCAATCCA GAGGAATAC CCATGTTTAC AGCAATTATCA GAAGGAGCCA
 1321 CCCCCACAAGA TTTAAACACA ATGTTAAATA CAGTGGGGGG ACATCAAGCA GCCATGCAA
 1381 TGTTAAAAGA TACCATCAAT GAGGAGGCTG CAGAAATGGG TAGGACACAT CCAGTACATG
 1441 CAGGGCTGT TGCACCAGGC CAGATGAGAG AACCAAGGGG AAGTGCACATA GCAGGAAC
 1501 CTAGTACCCCT TCAGGAACAA ATAGCATGGA TGACAAGTAA TCCACCTATT CCAGTAAAG
 1561 ACATCTATAA AAGATGGATA ATTCTGGGGT TAAATAAAAT AGTAAGAATG TATAGCCCTG
 1621 TTAGCATTTT GGACATAAAA CAAGGGCCAA AAGAACCCCTT TAGAGACTAT GTAGACCGGT
 1681 TCTTIAAAAC CTTAAGAGCT GAACAAGCTA CACAAGATGT AAAGAATTGG ATGACAGACA
 1741 CTTTGTGGT CAAAATGCG AACCCAGATT GTAAGACCAT TTTAAGAGCA TTAGGAC
 1801 GGGCCTCATT AGAAGAAATG ATGACAGCAT GTCAGGGAGT GGGAGGACT AGCCTAAAG
 1861 CAAGAGTGT GGCTGAGGCA ATGAGCCAAG CAAACAGTAA CATACTAGTG CAGAGAAC
 1921 ATTAAAAGG CTCAACAGA ATTATTAAT GTTCAACTG TGGCAAAGTA GGGCACATAG
 1981 CCAGAAATTG CAGGGCCCT AGGAAAAGG GCTGTTGGAA ATGTGGACAG GAAGGACACC
 2041 AAATGAAAGA CTGTAATGAG AGGCAGGCTA ATTITTTAGG GAAAATTGG CCTTCCAC
 2101 AGGGGAGGCC AGGGAATTTC CTCCAGAACCA GACCAGAGCC AACAGCCCCA CCAGCAGAAC
 2161 CAACAGCCCC ACCAGCAGAG AGCTTCAGGT TCGAGGAGAC AACCCCCGTG CCGAGGAAGG
 2221 AGAAAGAGAG GGAACCTITA ACTTCCCTCA AATCACTCTT TGGCAGGGAC CCCTTGTCTC
 2281 AATAAAAGTA GAGGGCCAGA TAAAGGAGGC TCTCTTAGAC ACAGGAGCAG ATGATACAGT
 2341 ATTAGAAGAA ATAGATTTGC CAGGGAAATG GAAACAAAA ATGATAGGGG GAATTGGAGG
 2401 TTTTATCAA GTAAAGACAGT ATGATCAAAT ACTTATAGAA ATTGTGGAA AAAAGGCTAT
 2461 AGGTACAGTA TTAGTAGGGC CTACACCACT CAACATAATT GGAAGAAATC TGTAACTCA
 2521 GCTTGGATGC ACACATAATT TTCCAATTAG TCCTATTGAA ACTGTACCG TAAAATTAAA
 2581 ACCAGGAATG GATGGCCCAA AGGTCAAACA ATGCCATTG ACAGAAGAAA AAATAAAAGC
 2641 ATTAACAGCA ATTGTGAGG AAATGGAGAA GGAAGGAAAA ATTACAAAAA TTGGCCTGA
 2701 TAATCCATAT AACACTCCAG TATTTGCCAT AAAAAAGAAG GACAGTACTA AGTGGAGAAA
 2761 ATTAGTAGAT TTCAGGGAAC TCAATAAAAG AACTCAAGAC TTTTGGGAAG TTCAATTAGG
 2821 AATACCACAC CCAGCAGGAT TAAAAAGAA AAAATCAGTG ACAGTGCTAG ATGTGGGGGA
 2881 TGCATATTTC TAGATGAAAG CTTCAAGGAAA TATACTGGCAT TCACCATACC

FIGURE 11

2941 TAGTATAAAC AATGAAACAC CAGGGATTAG ATATCAATAT AATGTGCTGC CACAGGGATG
 3001 GAAAGGATCA CCAGCAATAT TCCAGAGTAG CATGACAAAAA ATCTTAGAGC CCTTCAGAGC
 3061 AAAAAATCCA GACATAGTTA TCTATCAATA TATGGATGAC TTGTATGTAG GATCTGACTT
 3121 AGAAATAGGG CAACATAGAG CAAAATAGA AGAGTTAAGG GAACATTTAT TGAAATGGGG
 3181 ATTTACAACA CCAGACAAGA AACATCAAAA AGAACCCCCA TTTCTTGGA TGGGGTATGA
 3241 ACTCCATCCT GACAAATGGA CAGTACAACC TATACTGCTG CCAGAAAAGG ATAGTTGGAC
 3301 TGTCATGAT ATACAGAAGT TAGTGGAAA ATTAAACTGG GCAAGTCAGA TTACCCAGG
 3361 GATTAAGTA AGGCAACTCT GTAAACTCCT CAGGGGGGCC AAAGCACTAA CAGACATAGT
 3421 ACCACTAATC GAAGAAGCAG AATTAGAATT GGCAGAGAAC AGGGAAATT TAAGAGAAC
 3481 AGTACATGGA GTATATTATG ATCCATCAA AGACTTGATA GCTGAAATAC AGAAACAGGG
 3541 GCATGAACAA TGGACATATC AAATTATCA AGAACCATTT AAAATCTGA AAACAGGGAA
 3601 GTATGCAAAA ATGAGGACTA CCCACACTAA TGATGTAAAA CAGTTAACAG AGGCAGTGCA
 3661 AAAAAATAGCC ATGGAAAGCA TAGTAATATG GGGAAAGACT CCTAAATTAA GACTACCCAT
 3721 CCAAAAGAA ACATGGGAGA CATGGTGGAC AGACTATTGG CAAGCCACCT GGATCCCTGA
 3781 GTGGGAGTTT GTTAATACCC CTCCCTAGT AAAATTATGG TACCAACTAG AAAAGAGATCC
 3841 CATAGCAGGA GTAGAAACTT TCTATGTAGA TGGAGCAACT AATAGGGAAAG CTAAAATAGG
 3901 AAAAGCAGGG TATGTTACTG ACAGAGGAAG GCAGAAAATT GTTACTCTAA CTAACACAAAC
 3961 AAATCAGAAG ACTGAGTTAC AAGCAATTCA GCTAGCTCTG CAGGATTCAAG GATCAGAAGT
 4021 AAACATAGTA ACAGACTCAC AGTATGCATT AGGAATCATT CAAGCACAAC CAGATAAGAG
 4081 TGACTCAGAG ATATTTAACC AAATAATAGA ACAGTTAATA AACAAGGAAA GAATCTACCT
 4141 GTCATGGGTA CCAGCACATA AAGGAATTGG GGGAAATGAA CAAGTAGATA AATTAGTAAG
 4201 TAAGGGAAATT AGGAAAGTGT TGTTCTAGA TGGAAATAGAT AAAGCTCAAG AAGAGCATGA
 4261 AAGGTACAC AGCAATTGGA GAGCAATGGC TAATGAGTTT AATCTGCCAC CCATAGTAGC
 4321 AAAAGAAATA GTAGCTAGCT GTGATAAATG TCAGCTAAA GGGGAAGCCA TACATGGACA
 4381 AGTCGACTGT AGTCCAGGG AATGGCAATT AGATTGTACC CATTAGAGG GAAAATCAT
 4441 CCTGGTAGCA GTCCATGTAG CTAGTGGCTA CATGGAAGCA GAGGTATCC CAGCAGAAAC
 4501 AGGACAAGAA ACAGCATATT TTATTTAAA ATTAGCAGGA AGATGCCAG TCAAAGTAAT
 4561 ACATACAGAC AATGGCAGTA ATTTACCAAG TACTGCAGTT AAGGCAGCCT GTGGTGGGC
 4621 AGGTATCCAA CAGGAATTG GAATTCCCTA CAATCCCCAA AGTCAGGGAG TGTTAGAAC
 4681 CATGAATAAA GAATTAAGA AAATAATAGG ACAAGTAAGA GATCAAGCTG AGCACCTTAA
 4741 GACAGCAGTA CAAATGGCAG TATTCA CAAATTAAA AGAAAAGGGG GAATTGGGG
 4801 GTACAGTGCA GGGGAAAGAA TAATAGACAT AATAGCAACA GACATACAAA CTAAGAATT
 4861 ACAAAACAA ATTATAAGAA TTCAAAATT TCGGGTTTAT TACAGAGACA GCAGAGACCC
 4921 TATTGGAAA GGACCAGCCG AACTACTCTG GAAAGGTGAA GGGTAGTAG TAATAGAAGA
 4981 TAAAGGTGAC ATAAAGGTAG TACCAAGGAG GAAAGCAAAA ATCATTAGAG ATTATGGAAA
 5041 ACAGATGGCA GGTGCTGATT GTGTGGCAGG TGGACAGGAT GAAGATTAGA GCATGGAATA
 5101 GTTTAGTAAA GCACCATATG TATATATCAA GGAGAGCTAG TGGATGGTC TACAGACATC
 5161 ATTTGAAAG CAGACATCCA AAAGTAAGTT CAGAAGTACA TATCCATTAA GGGATGCTA
 5221 GATTAGTAAT AAAACATAT TGGGGTTTGC AGACAGGAGA AAGAGATTGG CATTGGGTC
 5281 ATGGAGTCTC CATAGAATGG AGACTGAGAG AATACAGCAC ACAAGTAGAC CCTGACCTGG
 5341 CAGACCAGCT AATTACATG CATTATTTTG ATTGTTTAC AGAATCTGCC ATAAGACAAG
 5401 CCATATTAGG ACACATAGTT TTTCCTAGGT GTGACTATCA AGCAGGACAT AAGAAGGTAG
 5461 GATCTCTGCA ATACTGGCA CTGACAGCAT TGATAAAACC AAAAAAGAGA AAGCCACCTC
 5521 TGCCTAGTGT TAGAAAATT GTAGAGGATA GATGGAACGA CCCCCAGAAG ACCAGGGGCC
 5581 GCAGAGGGAA CCATACAATG AATGGACACT AGAGATTCTA GAAGAACTCA AGCAGGAAGC
 5641 TGTCAGACAC TTTCCTAGAC CATGGCTCCA TAGCTTAGGA CAATATATCT ATGAAACCTA
 5701 TGGGGATACT TGGACGGGAG TTGAAGCTAT AATAAGAGTA CTGCAACAAAC TACTGTTCAT
 5761 TCATTCAGA ATTGGATGCC AACATAGCAG AATAGGCATC TTGCGACAGA GAAGAGCAAG
 5821 AAATGGAGCC AGTAGATCCT AAACTAAAGC CCTGGAACCA TCCAGGAAGC CAACCTAAA
 5881 CAGCTTGTAA TAATTGCTTT TGCAAACACT GTAGCTATCA TTGCTAGTT TGCTTICAGA

FIGURE 11

5941 CAAAAGGTTT AGGCATTTC TATGGCAGGA AGAAGCGGAG ACAGCGACCA AGCGCTCGTC
 6001 CAAGTGGTGA AGATCATCAA AATCCCTAT CAAAGCAGTA AGTACACATA GTAGATGTAA
 6061 TGGTAAGTTT AAGTTTATTT AAAGGAGTAG ATTATAGATT AGGAGTAGGA GCATTGATAG
 6121 TAGCACTAAT CATAGCAATA ATAGTGTGGA CCATAGCATA TATAGAATAT AGGAAATTGG
 6181 TAAGACAAAA GAAAATAGAC TGGTTAATTAAAGAATTAG GGAAAGAGCA GAAGACAGTG
 6241 GCAATGAGAG TGATGGGGAC ACAGAAGAAT TGTCAACAAT GGTGGATATG GGGCATCTTA
 6301 GGCTTCTGGA TGCTAATGAT TTGTAACACG GAGGACTTGT GGGTCACAGT CTACTATGGG
 6361 GTACCTGTGT GGAGAGAAC AAAAACTACT CTATTCTGTG CATCAGATGC TAAAGCATAT
 6421 GAGACAGAAC TGCTAATGTTCTGGCTACA CATGCTTGTG TACCCACAGA CCCCCAACCCA
 6481 CAAGAAATAG TTTTGGAAA TGTAACAGAA AATTAAATA TGTGGAAAAAA TAACATGGCA
 6541 GATCAGATGC ATGAGGATAT AATCAGTTA TGGGATCAAA GCCTAAAGCC ATGTGTAAAG
 6601 TTGACCCAC TCTGTGTAC TTTAAACTGT ACAGATAACAA ATGTTACAGG TAATAGAACT
 6661 GTTACAGGTA ATACAAATGA TACCAATATT GCAAATGCTA CATATAAGTA TGAAGAAATG
 6721 AAAAATTGCT CTTCAATGC AACACAGAA TTAAGAGATA AGAAACATAA AGAGTATGCA
 6781 CTCTTTATA AACTTGATAT AGTACCACTT AATGAAAATA GTAACAACCTT TACATATAGA
 6841 TTAATAAATT GCAATAACCTC AACCATACA CAAGCCTGTC CAAAGGTCTC TTTTGACCCG
 6901 ATTCCCTATAC ATTACTGTGC TCCAGCTGAT TATGCGATTC TAAAGTGTAA TAATAAGACA
 6961 TTCAATGGGA CAGGACCATG TTATAATGTC AGCACAGTAC AATGTACACA TGGAATTAAG
 7021 CCAGTGGTAT CAACTCAACT ACTGTTAAAT GGTAGTCTAG CAGAAGAACGG GATAATAATT
 7081 AGATCTGAAA ATTGACAGA GAATACAAA ACAATAATAG TACATCTTAA TGAATCTGTA
 7141 GAGATTAATT GTACAAGGCC CAACAATAAT ACAAGGAAAAA GTGTAAGGAT AGGACCAGGA
 7201 CAAGCATTCT ATGCAACAAA TGACGTAATA GGAAACATAA GACAAGCACA TTGTAACATT
 7261 AGTACAGATA GATGGAATAA AACTTACAA CAGGTAATGA AAAAATTAGG AGAGCATTTC
 7321 CCTAATAAAA CAATAAAATT TGAACCACAT GCAGGAGGGG ATCTAGAAAT TACAATGCAT
 7381 AGCTTTAATT GTAGAGGAGA ATTTCCTAT TGCAATACAT CAAACCTGTT TAATAGTACA
 7441 TACTACCCCTA AGAATGGTAC ATACAAATAC AATGTAATT CAAGCTTACCC CATCACACTC
 7501 CAATGCAAAA TAAAACAAAT TGTACGCTAT TGGAAGGGG TAGGACAAGC AATGTATGCC
 7561 CCTCCCATTG CAGGAAACAT AACATGTAGA TCAAACATCA CAGGAATACT ATTGACACGT
 7621 GATGGGGGAT TTAACAACAC AAACAACGAC ACAGAGGAGA CATTGAGACC TGGAGGAGGA
 7681 GATATGAGGG ATAACCTGGAG AAGTGAATTAA TATAAATATA AAGTGGTAGA AATTAAGCCA
 7741 TTGGGAATAG CACCCACTAA GGCAAAAAAGA AGAGTGGTGC AGAGAAAAAA AAGAGCAGTG
 7801 GGAATAGGAG CTGTGTTCTT TGGGTTCTTG GGAGCAGCAG GAAGCACTAT GGGCGCAGCG
 7861 TCAATAACGC TGACGGTACA GCCCAGACAA CTGTTGCTG GTATAGTGCACAGC
 7921 AATTGCTGA AGGCTATAGA GGCGCAACAG CATATGTTGC AACTCACAGT CTGGGGCATT
 7981 AAGCAGCTCC AGGGCAGAGT CCTGGCTATA GAAAGATACC TAAAGGATCA ACAGCTCCTA
 8041 GGGATTTGGG GCTGCTCTGG AAGACTCATC TGCAACACTG CTGTGCTTGT GAACTCCAGT
 8101 TGGAGTAATA AATCTGAAGC AGATATTGAG GATAACATGA CTTGGATGCA GTGGGATAGA
 8161 GAAATTAAATA ATTACACAGA AACAAATATTC AGGTTGCTTG AAGACTCGCA AAACCAGCAG
 8221 GAAAAGAAATG AAAAAGATTT ATTAGAAATG GACAAGTGGAA ATAATCTGTG GAATTGGTTT
 8281 GACATATCAA ACTGGCTGTG GTATATAAAA ATATTCAAA TGATAGTAGG AGGCTTGATA
 8341 GGTTTAAGAA TAATTTTGTG TGTGCTCTCT ATAGTGAATA GAGTTAGGCA GGGATACTCA
 8401 CCTTTGTCAT TTCAGACCCCT TACCCCAAGC CCGAGGGGAC TCGACAGGCT CGGAGGAATC
 8461 GAAGAAGAAC GTGGAGAGCA AGACAGAGAC AGATCCATAC GATTGGTGAG CGGATTCTTG
 8521 TCGCTTGCTT GGGACGATCT GCGGAGCTG TGCCCTTCA GCTACCACCG CTTGAGAGAC
 8581 TTCATATTAA TTGCACTGTG GGCAGTGAA CTTCTGGGAC ACAGCAGTCT CAGGGGACTA
 8641 CAGAGGGGTTT GGGAGATCCT TAAGTATCTG GGAAGTCTTG TGCAGTATTG GGGTCTAGAG
 8701 CTAAAAAAGA GTGCTATTAG TCCGCTTGAT ACCATAGCAA TAGCACTAGC TGAAGGAACA
 8761 GATAGGATTA TAGAATTGGT ACAAAAGAATT TGTAGAGCTA TCCTCAACAT ACCTAGGAGA
 8821 ATAAGACAGG GCTTGAAGC AGCTTGTCA TAAAATGGGA GGCAAGTGGT CAAAACGCAG
 8881 CATAGTTGGA TGGCCTGCAG TAAGAGAAC AATGAGAACAG ACTGAGCCAG CAGCAGAGGG
 8941 AGTAGGAGCA GCGTCTCAAG ACTTAGATAG ACATGGGGCA CTTACAAGCA GCAACACACC

FIGURE 11

9001 TGCTACTAAT GAAGCTTGTG CCTGGCTGCA AGCACAAAGAG GAGGACGGAG ATGTAGGCTT
9061 TCCAGTCAGA CCTCAGGTAC CTTAACGACC AATGACTTAT AAGAGTCAG TAGATCTCAG
9121 CTTCTTTTA AAAGAAAAGG GGGGACTGGA AGGGTTAATT TACTCTAGGA AAAGGCAAGA
9181 AATCCTTGAT TTGTGGGTCT ATAACACACA AGGCTTCTTC CCTGATTGGC AAAACTACAC
9241 ATCGGGGCCA GGGGTCCGAT TCCCAGTGAC CTTTGGATGG TGCTTCAAGC TAGTACCACT
9301 TGACCCAAGG GAGGTGAAAG AGGCCAATGA AGGAGAAGAC AACTGTTTGC TACACCCAT
9361 GAGCCAACAT GGAGCAGAGG ATGAAGATAG AGAAGTATTA AAGTGGAAAGT TTGACAGCCT
9421 TCTAGCACAC AGACACATGG CCCCGAGCT ACATCCGGAG TATTACAAAG ACTGCTGACA
9481 CAGAAGGGAC TTTCCGCCTG GGACTTTCCA CTGGGGCGTT CCGGGAGGTG TGGTCTGGGC
9541 GGGACTTGGG AGTGGTCACC CTCAGATGCT GCATATAAGC AGCTGCTTTT CGCTTGTACT
9601 GGGTCTCTCT CGGTAGACCA GATCTGAGCC TGGGAGCTCT CTGGCTATCT AGGAAACCCA
9661 CTGCTTAGGC CTCATAAAAG CTTGCCTTGA GTGCTCTAAG TAGTGTGTGC CCATCTGTTG
9721 TGTGACTCTG GTAATAGAG ATCCCTCAGA CCCTTGTGG TAGTGTGGAA AATCTCTAGC
9781 A

FIGURE 11

SEQ ID NO:34

GCTGAGGCAATGAGCCAAGCAACCAGCGCAAACATACTGATGCAGAGAAGCAATT
CAAAGGCCCTAAAAGAATTATTAAATGTTCAACTGTGGCAAGGAAGGGCACATAG
CTAGAAATTGTAGGGCCCCTAGGAAAAAAGGCTGTTGGAAATGTGGAAAGGAAGGA
CACCAAATGAAAGACTGTACTGAGAGGCAGGCTAA

FIGURE 12

975Pol wt until 6aa Int: (SEQ ID NO:35)

TTTTTAGGGAGATTGGCTTCCACAAGGGAAAGGCCAGGGAATTCCTTCAGAA
CAGAACAGAGCCAACAGCCCCACCAGCAGAGAGCTCAAGTCAGGAGACAAACCC
CCGCTCCGAAGCAGGAGCGAAAGACAGGGAACCCCTAATTCCCTCAAATCACTCT
TTGGCAGCGACCCCTGTCTCAATAAAAGTAGGGGTCAAATAAAGGAGGCTCTT
AGACACAGGAGCTGATGATACAGTATTAGAAGAAATGAGTTGCCAGGAAAATGGA
AACCAAAAATGATAGGAGGAATTGGAGGTTTATCAAAGTAAGACAGTATGATCAA
ATACTTATAGAAATTGAGAAAAAGGCTATAGGTACAGTATTAAATAGGACCTACA
CCTGTCAACATAATTGAGAAATATGTTGACTCAGCTGGATGCACACTAAATT
CCAATTAGTCCCATTGAAACTGTGCCAGTAAATTAAAGCCAGGAATGGATGGCCA
AAGGTTAAACAATGGCCATTGACAGAAGAGAAAATAAAGCATTAAACAGCAATTG
TGAAGAAATGGAGAAAGAAGGAAAATTACAAAAATTGGGCTGAAAATCCATATA
ACACTCCAGTATTGCCATAAAAAGAAGGACAGTACTAAGTGGAGAAAGTTAGTA
GATTTCAGGGAACTTAATAAAAGAACTCAAGACTTTGGAGTTCAATTAGGAATA
CCACACCCAGCAGGGTTAAAAAAGAAAAATTCACTGACAGTACTGGATGTGGGGA
TGCATATTITCAGITCCTTGTAGATGAGGACTTCAGGAAATATACTGCATTACCCATA
CCTAGTATAAACATGAAACACCAGGGATTAGATATCAATATAATGTGCTTCCACAG
GGATGGAAAGGATCACCATCAATATTCCAGAGTAGCATGACAAAAATTAGAGCC
CTTAGAGCAAGAAATCCAGAAATAGTCATCTATCAATATAATGGATGACTTGTATGT
AGGATCTGACTTAGAAATAGGGCAACATAGAGCAAAAATAGAGGAGTTAAGAAAAC
ATCTGTTAAGGTGGGGATTACCACACCGGACAAGAAACATCAGAAAGAACCCCCA
TTCTTGGATGGGTATGAACTCCATCCTGACAAATGGACAGTACAGCCTATAGAG
TTGCCAGAAAAGGAAAGCTGGACTGTCAATGATATACAGAAGTTAGTGGAAAATT
AAATTGGGCCAGTCAGATTACCCAGGAATTAAAGTAAGGCAACTTGTAAACTCCT
TAGGGGGGCCAAAGCACTAACAGATATACTGACACTAATGGAGTATATTAG
AATTGGCAGAGAACAGGGAAATTCTAACAGAGAACACAGGGCATGACAAATGGACATA
CCATCAAAAGACTTGGTAGCTGAAATACAGAACACAGGGAGTATGCAAAAATGA
TCAAATTACCAAGAACCAATTCAAAACCTGAAACACAGGGAGTATGCAAAAATGA
GGACTGCCACACTAACAGTTAACAGAGGAGTGCACAGGAAATAGCT
ATGGAAAGCATAGTAATATGGGAAAGACTCCTAAATTAGACTACCCATCCAAA
AGAAACATGGAGACATGGTAGCAGACTATTGGCAAGCCACCTGGATTCTGAGT
GGGAGTTGTTAACACCTCCCTAGTAAATTATGGTACCAAGCTAGAGAAAGAAC
CCATAATAGGAGCAGAAACTTCTATGTAGATGGAGCAGCTAACAGGGAAACTAAA
ATAGGAAAAGCAGGGTATGTACTGACAGAGGAAGGCAGAAAATTGTTCTCTAAC
AGAAACAAACAAATCAGAAGACTGAATTACAAGCAATTCACTGCTAGCTTGCAAGATT
AGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAG
CACAACCAAGATAAGAGTGAATCAGAGTTAGTCACCAACAAATAATAGAACAAATTAAATA
AAAAAGGAAAAGGTCTACCTGTCATGGGTACCAAGCACATAAAGGAATTGGAGGAAA
TGAACAAATAGATAAATTAGTAAGTAAGGAAATCAGGAAAGTGTGCTTTCTAGATG
GAATAGAT

FIGURE 13

SEQ ID NO:36

GGCGGCATCGTGATCTACCAAGTACATGGACGACCTGTACGTGGGCAGCGGCG
GC

FIGURE 14

WO 02/04493

PCT/US01/21241

18/114

SEQ ID NO: 37

GGIVIYQYMDDLYVGSGG

FIGURE 15

12_5/1ZA (SEQ ID NO:45)

TGGAAGGGTTAATTACTCCAGGAAAAGGCAAGAGATCCTGATTATGGGTCTATC
ACACACAAGGCTACTTCCCTGATTGGCAAAACTACACACCAGGGACCAGGGTCAGA
TATCCACTGACCTTGGATGGTGCTCAAGCTAGTGCCAGTGACCCAAGGGAAGTA
GAAGAGGCCAACGGAGGAGAAGACAACAGTGTACAGTGGAAAGTTGACAGCAGCTAGCAC
AATGGATGATGAACACAAAGAAGTGTACAGTGGAAAGTTGACAGCAGCTAGCAC
GCAGACACCTGGCCCGCGAGCTACATCCGGATTATTACAAAGACTGCTGAACAGA
AGGGACTTCCGCTGGGACTTCCACTGGGGCGTCCAGGGGAGTGGTCTGGCG
GGACTGGGAGTGGCCAGCCCTCAGATGCTGCATAAGCAGCGGCTTTCGCTGTA
CTGGGTCTCTCTAGGTAGACAGATCCGAGCCTGGGAGCTCTGTCTATCTGGGA
ACCCACTGCTTAGGCCTCAATAAAAGCTTGCCTTGAGTGCTCAAGTAGTGTGCCC
ATCTGTTGTGACTCTGGTAACCTGGTAACTAGAGATCCCTCAGACCCCTTGTGGT
AGTGTGGAAAATCTCTAGCAGTGGCGCCGAACAGGGACTGAAAGCGAAAGTGAG
ACCAGAGAAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGTGCACTCGGCAAGAG
GCGAGGGGGCGACTGGTAGTACGCCAAAATTTTTACTAGCGGAGGCTAGA
AGGAGAGAGATGGGTGCGAGAGCGTCAATATTAAGAGGGGAAATAGACAAAT
GGGAAAAAAATTAGGTACGCCAGGGGGAGAAAACACTATATGCTAAAACACCTA
GTATGGCAAGCAGAGAGCTGAAAGATTGCAGTTAACCTGGCTTTAGAGAC
ATCAGACGGATGTAGAC AAATAATAAAACAGCTACAACCAGCTTCA
CAGGAACAGAGGAAATTAGATCATTATTAAACACAGTAGCAACTCTCTATTGTGTAC
ATAAAGGGATAGATGTACGAGACACCAAGGAAGCCTTAGACAAGATAGAGGAGGA
ACAAAAACAAATGTCAGAAAAAACACAGCAGGCCAGCGCTGACAAAAAGGTC
AGTCAAAATTATCCTATAGTCAGAACCTCCAAGGGCAAATGGTACACCAGGCCAT
ATCACCTAGAACCTTGAATGCATGGTAAAAGTAATAGAGGAGAAGGCTTGT
CAGAGGTAATACCCATGTTACAGCATTATCAGAAGGAGCCACCCACAAGATT
AACACCATGTTAAATACAGTGGGGGACATCAAGCAGCCATGCAAATGTTAAAG
ATACCATCAATGAGGAGGCTGCAGAATGGTAGGTACATCCAGTACATGCAGGG
CCTGTTGACCAGGCCAGATGAGAGAACCAAGGGAAAGTGACATAGCAGGAAC
CTAGTACCCCTCAAGAACAAATAGCATGGATGACAAGTAACCCACCTATCCCAGTA
GGGGACATCTATAAAAGGTGGATAATTCTGGGTTAAATAAAATAGTAAGAATGTA
CAGCCCTGTCAAGCATTAGACATAAAACAAGGACCAAGGAACCCCTTAGAGACT
ATGTAGACCGGTTCTCAAAACTTAAAGAGCTGAACAATCTACACAAGAGGTAAA
AATTGGATGACAGACACCTTGTAGTCCAAAATGCGAACCCAGATTGTAAGACCATT
TTAAGAGCATTAGGACCAGGGCTCATTAGAAGAAATGATGACAGCATGTCAGGG
AGTGGGAGGACCTAGCCACAAAGCAAGAGTTGGCTGAGGCAATGAGCAAGCAA
ACAATACAAGTGTAAATGATAAGAAAAGCAATTAAAGGCCCTAGAAGAGCTGTT
AAATGTTCAACTGTGGCAGGGAAAGGGCACATGCCAGGAATTGCAGGGCCCTAG
GAAAAGGGCTGTGGAAATGTGGAAAGGAAGGACACCAATGAAAGACTGTACT
GAGAGGCAGGCTAATTAGGAAATTTGGCCTCCACAAGGGAGGCCAGG
GAATTCCCTCAGAGCAGACAGCCACAGCCCCACCACTAGAACCAACAGCCC
CACCAGCAGAGAGCTCAAGTCAAGGAGACTCCGAAGCAGGAGCCGAAAGACAG
GGAACCTTAACTCCCTCAAATCACTTTGGCAGCGACCCCTGTCTCAATAAAA

FIGURE 16

GTAGCGGGCAAACAAAGGAGGCTTTAGATACAGGAGCAGATGATACAGTACT
AGAAGAAAATAAACTGCCAGGAAAATGGAAACCAAAATGATAGGAGGAATTGGA
GGTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTGTGGAAAAGG
GCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTG
TTGACTCAGCTGGATGCACACTAAATTCCAATTAGCCCCATTGAAACTGTACCA
GTAAAATTAAAGCCAGGAATGGATGGCCAAAGGTTAACAAATGGCCATTGACAGA
AGAAAAAATAAAAGCATTAAACAGAAATTGTGAGGAATGGAGAAGGAAGGAAAAA
ATTACAAAATTGGGCCTGAAAATCCATATAACACTCCAGTATTGCCATAAAGAAC
AAGGACAGTACAAAGTGGAGAAATTAGTAGATTTCAGGGAACCTCAATAAAAGAAC
TCAAGACTTTGGGAAGTCCAATTAGGAATACCACACCCAGCAGGGTAAAGAAGA
AAAAATCAGTGACAGTACTGGATGTGGAGATGCATATTTCAGTCCCTTAGATG
AGAGCTTCAGAAAATATACTGCATTACCATACCTAGTATAAACAAATGAAACACCA
GGGATTAGATATCAATATAATGTTCTCACAGGGATGGAAAGGATCACCAGCAA
TATTCCAGAGTAGCATGACAAGAATCTAGAGGCCCTTAGAACACAAAACCCAGAA
GTAGTTATCTATCAATATATGGATGACTTATATGTAGGATCTGACTTAGAAATAGGG
CAACATAGAGCAAAATAGAGGAGTTAACAGGACACCTATTGAAATGGGGATTAC
CACACCAGACAAGAAACATCAGAAAGAACCCCCATTCTTGATGGGTATGAAC
TCCATCCTGACAAATGGACAGTACAGCCTATACAGCTGCCAGAAAAGGAGAGCTGG
ACTGTCAATGATATACAGAAGTTAGGGAAAGTTAACACTGGCAAGTCAGATTAA
CCCAGGGATTAAAGTAAGGCAACTGTGTAAACTCCTTAGGGAGCCAAAGCACTAA
CAGACATAGTCCACTGACTGAAGAACAGCAGAATTAGAACATTGGCTGAGAACAGGGA
AATTCTAAAAGAACCAAGTACATGGAGTATATTGACCCATCAAAAGATTAAATAG
CTGAAATACAGAAACAGGGGAATGACCAATGGACATATCAAATTACCAAGAAC
ATTAAAAATCTGAGAACAGGAAAGTATGCAAAATGAGGACTGCCACACTAATG
ATGTGAAACAGTTAGCAGAGGCAGTGCAAAAGATAACCCAGGAAAGCAGATAGTAATA
TGGGGAAAAACTCCTAAATTAGACTACCCATCCAAAAGAAACATGGGAGACATG
GTGGTCAGACTATTGGCAAGCCACCTGGATTCTGAGTGGAGTTGTCAATACCC
TCCCCTAGAAAATTGTGGTACCAAGCTGGAAAAAGAACCCATAGTAGGGCAGAAA
CTTCTATGTAGATGGAGCAGCCAATAGGGAAACTAAAATAGGAAAGCAGGGTAT
GTCACTGACAAGGAAGGCAGAAAGTTGTTCTCACTGAAACAAACAATCAGAA
GACTGAATTACAAGCAATTCACTAGCTAGCTTGAGGATTAGGGCCAGAAGTAAACA
TAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAACCAAGATAAGAGT
GAATCAGAATTAGTCAGTCAAATAATAGAACAGTTGATAAAAAAGGAAAAAGTCTA
CCTATCATGGTACCAAGCACATAAAGGAATTGGAGGAAATGAACAAGTAGACAAAT
TAGTAAGTAGTGGAAATCAGAAAAGTACTGTTCTAGATGGAATAGATAAAGCTCAA
GAAGAGCATGAAAATATCACAGCAATTGGAGAGCAATGGTAGTGGTTAATCT
GCCACCCATAGTAGCAAAGGAAATAGTAGCCAGCTGTGATAAAATGTCAGCTAAAAG
GGGAAGCCATGCATGGACAAGTCGACTGTAGTCCAGGAATATGGCAATTAGACTGT
ACACATTAGAAGGAAAATCATCCTAGTAGCAGTCCATGTAGCCAGTGGCTACAT
GGAAGCAGAGGTTATCCACAGCAGAAACAGGACAAGAACAGCATACTTATACTAA
AATTAGCAGGAAGATGGCCAGTCAAAGTAATACATACAGATAATGGCAGTAATTTC
ACCAGTACCGCAGTTAAGGCAGCCTGTTGGGGCAGATATCCAACGGGAATTG
AATTCCCTACAATCCCCAAAGTCAGGAGTAGTGAATCCATGAATAAAGAATTAA

FIGURE 16

AGAAAATCATAGGGCAAGTAAGAGATCAAGCTGAGCACCTTAAGACAGCAGTACAA
ATGGCAGTATTCAATTACAATTAAAAGAAAAGGGGGATTGGGGGTACAGTGC
AGGGGAGAGAATAATAGACATAATAGCATCAGACATACAAACTAAAGAATTACAAA
AACAAATTATAAAATTCAAATTTCGGGTTATTACAGAGACAGCAGAGACCTTA
TTGGAAAGGACCAGCCAAACTACTCTGGAAAGGTGAAGGGCAGTAGTAATACAA
GATAATAGTGTATATAAGGTAGTACCAAGAAGGAAAGCAAAATCTAAGGACTA
TGGAAAACAGATGGCAGGTGCTGATTGTGTCAGGTAGACAGGATGAAGGATTAGA
ACATGGCAGCTTAGTAAGCACCATACTGTATGTTCGAGGAGAGCTGATGGATGG
TTCTACAGACATCATTGAAAGCAGACACCCAAAAGTAAGTCAGAAGTACACAT
CCCATTAGGAGATGCCAGGTTAGTAATAAAACATAATTGGGCTGCAGACAGGAG
AAAGAGCTTGGCATTGGTCACGGAGTCTCCATAGAATGGAGATTGAGAAGATAT
AGCACACAAAGTAGACCCCTGACCTGACAGACCAACTAATTCTATGCATTATTGAT
TGTTTGCGAGAATCTGCCATAAGGAAAGCCATACTAGGACAGATAGTTAGCCCTAA
GTGTGACTATCAAGCAGGACATAACAAGGTAGGATCTCTACAATACTTGGCACTGA
CAGCATTGATAAAACAAAAAGATAAAGCCACCTCTGCCTAGTGTAGGAAATTA
GTAGAGGATAGATGGAACAAGCCCCAGAAGACCAGGGGCCAGAGGAAACCATA
CAATGAATGGACACTAGAGCTTGAAGAAGACTCAAGCAGGAAGCTGTCAGACACT
TTCTAGACCATGGCTCCATAACTTAGGACAACATACTATGAAACCTATGGAGATA
CTTGGACAGGAGTGAAGCAATAAGAATCCTGCAACAATTACTGTTATTCTATT
TCAGGATTGGGTGCCATCATAGCAGAAATAGGCATTGCGACAGAGAAGAGCAAGA
AATGGAGCCAATAGATCCTAACCTAGAACCTGGAACCATCCAGGAAGTCAGCCTA
AAACTGCTTGTAAATGGGTGTTACTGTAAACGTTGCAGCTATCATTGCTAGTTGCTT
TCAGAAAAAAAGGCTAGGCATTACTATGGCAAGAAGACTGAGGAGACAGCGACGAA
GCGCTCCTCCAAGCAATAAGATCATCAAGATCCTCTACCAAAGCAGTAAGTACCG
AATAGTATATGTAATGTTAGATTAACTGCAAGAATAGATTCTAGATTAGGAATAGG
AGCATTGATAGTAGCACTAATCATAGCAATAATAGTGTGGACCATAGTATATAG
AATATAGGAAATTGGTAAGGAAAGGAAATAGACTGGTAGTTAAAAGGATTAGG
GAAAGAGCAGAACAGTGGCAATGAGAGCGAGGGGATACTGAAGAATTATCGA
CACTGGTGGATATGGGCATCTTAGGCTTGGATGCTAATGATGTAAATGTGAA
GGGCTTGTGGTCACAGTCTACTACGGGTACCTGTGGGAGAGAACGAAACT
ACTCTATTGTGCATCAGATGCTAAAGCATATGAGAAAGAAGTCATAATGTCTG
GGCTCACACATGCTGTGACCCACAGACCCACAAAGAAGTGTGATTGGGC
AATGTAACAGAAAATTAAACATGTGGAAAATGACATGGGATCAGATGCAGG
AAGATATAATCAGTTATGGGATCAAAGCCTAACGGCTAAGCCATGTGAAAATTGACCCCA
CTCTGTGTCACTTAAACTGTACAAATGCAACTGTTAACTACAATAACCTCTAAA
GACATGAAAATTGCTTCTATGTAACCACAGAAATTAGAGATAAGAAAAAGAA
AGAAAATGCACTTTATAGACTGTATAGTACCACTTAATAATAGGAAGAATGG
GAATATTAACAACATAGATAATAAAATTGTAATACCTCAGCCATAACACAAGCCTG
TCCAAAAGTCTGTTGACCCAACTCTATACATTATTGTGCTCCAGCTGGTTATGCG
CCTCTAAAATGTAATAATAAGAAATTCAATGGAATAGGACCATGCGATAATGTCAG
CACAGTACAATGTACACATGGAATTAAGCCAGTGGTATCAACTCAATTACTGTTAAA
TGGTAGCCTAGCAGAAGAAGAGATAATAATTAGATCTGAAAATCTGACAAACAATG
TCAAAACAATAATAGTACATCTTAATGAATCTATAGAGATTAAATGTACAAGACC

FIGURE 16

TGGCAATAATACAAGAAAGAGTGTGAGAATAGGACCAGGACAAGCATTCTATGCA
ACAGGAGACATAATAGGAGATATAAGACAAGCACATTGTAACATTAGTAAAAATGA
ATGGAATACAACATTACAAAGGGTAAGTCAAAATTACAAGAACACTCTCCCTAATA
GTACAGGGATAAAATTGCACCACACTCAGGAGGGGACCTAGAAATTACTACACAT
AGCTTAATTGTGGAGGAGAATTTCTATTGCAATACAACAGACCTGTTAATAGT
ACATACAGTAATGGTACATGCACTAATGGTACATGCATGTCTAATAATACAGAGCG
CATCACACTCCAATGCAGAATAAAACAAATTATAAACATGTGGCAGGAGGTAGGAC
GAGCAATGTATGCCCTCCCATTGCAAGGAAACATAACATGTAGATCAAATATTACA
GGACTACTATTAACACGTGATGGAGGAGATAATAACTGAAACAGAGACATTTCAG
ACCTGGAGGAGGAGACATGAGGGACAATTGGAGAAGTGAATTATAAAATACAAG
GTGGTAGAAATTAAACCATTAAGGAGTAGCACCCACTGCTGCAAAAGGAGAGTGGT
GGAGAGAGAAAAAGAGCAGTAGGAATAGGAGCTGTGTCCTGGTTCTGGGAG
CAGCAGGAAGCACTATGGCGCAGCATCAATAACGCTGACGGTACAGGCCAGACAA
TTATTGTCGGTATAGTCAACAGCAAAGTAATTGCTGAGGGCTATAGAGGCGCAA
CAGCATATGTCACACTACGGCTGGGCATTAAGCAGCTCCAGGCAAGAGTCCTG
GCTATAGAGAGATAACCTACAGGATCAACAGCTCTAGGACTGTGGGGCTGCTCTGG
AAAACTCATCTGCACCAACTATGTGCTTGGAACTCTAGTGGAGTAATAAAACTCA
AAGTGATATTGGGATAACATGACCTGGATGCAGTGGATAGGGAAATTAGTAATT
ACACAAACACAATATACAGGTTGCTTGAAGACTCGCAAAGCCAGCAGGAAAGAAA
TGAAAAAGATTACTAGCATGGACAGGTGGACAATCTGTGGAATTGGTTAGCAT
AACAAATTGGCTGGTATATAAAATATTCAATGATAGTAGGAGGCTTAGATAG
GTTAAGAATAATTGCTGTGCTCTCTAGTAAATAGAGTTAGGCAGGGATACT
CACCCCTGTCATTGCAGACCTTATCCAAACCCGAGGGGACCCGACAGGCTCGGA
GGAATCGAAGAAGAAGGTGGAGAGCAAGACAGCAGCAGATCCATTGATTAGTGA
GCGGATTCTGACACTTGCCTGGACGACCTACGAAGCCTGTGCCTCTGCTTACCC
ACCGATTGAGAGACTTCATATTAAATTGTTAGTGGAGAGCAGTGGAACTTCTGGGACAC
AGTAGTCTCAGGGACTGCAGAGGGGTGGGAACCCITAAGTATTGGGAGTCT
TGTGCAATATTGGGTCTAGAGTAAAAAGAGTGTATTAAATCTGCTGATACTAT
AGCAATAGCAGTAGCTGAAGGAACAGATAGGATTCTAGAATTCAACAAACCTT
GTAGAGGTATCCGCAACGTACCTAGAAGAATAAGACAGGGCTCGAAGCAGCTTG
CAATAAAATGGGGGCAAGTGGTAAAAAGCAGTATAATTGGATGGCTGAAGTAA
GAGAAAGAATCAGACGAACCTAGGTAGCAGCAGAGGGAGTAGGATCAGCGTCTCA
AGACTTAGAGAAACATGGGCACCTACAACCCAGCAACACAGCCCACAACATGCTG
CTTGCCTGGCTGGCAAGCGCAAGAGGGAGGAAGTAGGCTTCCAGTCAG
CCTCAGGTACCTTAAAGACCAATGACTTAAAGCAGCAATAGATCTCAGCTTCTT
TTAAAAGAAAAGGGGGACTGGAAGGGTTAATTACTCAAGAAAAGGCAAGAGAT
CCTGATTGTTGTTATAACACACAAGGCTTCTCCCTGATTGGCAAAACTACAC
ACCGGGACCAAGGGTCAGATTCCACTGACCTTGGATGGTACTCAAGCTAGAGCC
AGTCGATCCAAGGGAAAGTAGAAGAGGCCAATGAAGGAGAAAACAACACTGTTACTAC
ACCCTATGAGCCAGCATGGAATGGAGGATGAAGACAGAGAAGTATTAGATGGAAG
TTTGACAGTACGCTAGCACCGCAGACACATGGCCCGAGCTACATCCGGAGTATTAC
AAAGACTGCTGACACAGAAGGGACTTCCGCTGGGACTTCCACTGGGGCTTCCAG
GAGGTGTGGTCTGGCGGGACAGGGAGTAGGTCAGCCCTGAGATGCTGATATAAG
CAGCTGCTTTCGCTGTACTGGCTCTCTAGGTAGACCAGATCTGAGCCCCGGAG

FIGURE 16

CTCTCTGGCTATCTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTG
CCTTGAGTAGTGTGTGCCGTCTGTGTGACTCTGGTAAGAGATCCCTCAGA
CCACTTGTGGTAGTGTGGAAAATCTCTAGCA

FIGURE 16

>C4_Env_TV1_C_ZA_opt_short (SEQ ID NO: 46)

CATCACCCCTGCAGTGCAAGATCAAGCAGATCGTGCATGTGGCAGGGCGTGGGCCAGGCCATGTACGCCCCCCCCATCG
CCGGCAACATCACCTGC

FIGURE 17

>C4_Env_TV1_C_ZA_opt (SEQ ID NO:47)

CTGCCCATCACCTGCAGTGCAGATCAAGCAGATCGTGCATGTGGCAGGGCGTGGCCAGGCCATGTACCCCCCCC
CATGCCGGAACATCACCTGCCAGAACATCACCGGCATCCTGCTGACCCGCGACGGCGGC

FIGURE 18

>C4_Env_TV1_C_ZA_wt (SEQ ID NO:48)

TTACCCATCACACTCCAATGCAAAATAAAACAAATTGTACGCATGGCAAGGGTAGGACAAGCAATGTATGCCCTCC
CATTGCAGGAAACATAACATGTAGATCAAACATCACAGGAATACTATTGACACGTGATGGGGGA

FIGURE 19

>Envgp160_TV1_C_Zaopt (SEQ ID NO:49)

ATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGGTGGATCTGGGCATCCTGGCTTCTGGATGCTGATGATCTG
CAACACCGAGGGACCTGTGGGTGACCGTGTACTACGGCGTCCCCGTGTGGCGAGGCCAACACCCTGTTCTGCGCCA
GCGACGCCAAGGCCAACCGTGTGGGCAACACGAGGAGGTGCAACACGAGAAGAACCTGCGTGTGGCGACCGACCCCAACCCCAAG
GAGATCGTGTGGGCAACGTGACCGAGAACCTCAACATGTGGAAGAACAAACATGGCGACCAGATGCACGAGGACATCAT
CAGCGTGTGGGACCAAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTGTGGCGACCTGAACCTGACCCGACACCAACG
TGACCCGCAACCGCACCGTGAACCGGCAACACCAACGACACCAACATGCCAACGCCACCTACAAGTACGAGGAGATGAAG
AACTGCAGCTTCAACGGCACCCACCGAGCTGCCGACAAGAACGACAAGGAGTACGCCCTGTTCTACAAGCTGGACATCGT
GCCCTGAACGAGAACAGCAACAACCTCACCTACCGCCTGATCAACTGCAACACCAGCACCACCCAGGGCTGCCCCA
AGGTGAGCTTCGACCCCATTCCACTACTGCGCCCCCGCGACTACGCCATCTGAAGTGCAACAACAAGACCTTC
AACGGCACCGGCCCCCTGCTACAACGTGAGCACCGTGCAGTGACCCACGGCATCAAGCCCGTGGTGAAGCAACCCAGCTGCT
GCTGAACGGCAGGCTGGCGAGGAGGGCATCATCATCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
ACCTGAACGAGAGCGTGGAGATCAACTGCAACCCGCCCCAACAAACAACACCCGCAAGAGCGTGCATCGGCCCCGGCCAG
GCCCTTACGCCACCAACGACGTGATCGGCAACATCGGCCAGGCCACTGCAACATCAGCACCGACCGCTGGAACAAGAC
CTTGCAGCAGGTGATGAAGAAGCTGGCGAGCACTTCCCAACAGACCATCAAGTTGAGCCCCACGCCGGCGGCGACC
TGGAGATCACCATGCACAGCTCAACTGCCGGCGAGTCTTCTACTGCAACACCAAGCAACCTGTTCAACAGCACCTAC
TACCCCAAGAACGGCACCTACAAGTACAACGGCAACAGCAGCCTGCCCATCACCTGCAAGATCAAGCAGATCGT
GCGCATGTGGCAGGGCGTGGGCCAGGCCATGTCAGCCCCCCCCTGCGGCAACATCACCTGCCGAGCAACATCACCG
GCATCCCTGTGACCCCGCAGGGCGCTTCAACACACCAACAGACACCCGAGGAGACCTTCCGCCCCGGCGGCGAC
ATGCGGACAACACTGGCGCAGCGAGCTGATCAAGTACAAGTGGTGGAGATCAAGCCCTGGCATCGCCCCACCAAGGC
CAAGCGCCGCGTGGTGCAGCGAAGAAGCGCCGCGTGGCATCGGCCGCTGTTCTGGCTTCTGGCGCCGCG
GCACCATGGCGCCGCAGCATCACCTGACCGTGCAGGCCAGCTGCTGAGCGGCATCGCAGCAGCAGAGCAAC
CTGCTGAAGGCCATCGAGGCCAGCAGCACTGCTGAGCTGACCGTGTGGCATCAAGCAGCTGAGGCCGCGTGT
GCCATCGAGCGCTACCTGAAGGACCAAGCAGCAGCTGCTGGCATCTGGGCTGAGCGCCGCTGATCTGCACCACCGCCG
TGCCCTGGAACAGCAGCTGGAGCAACAAGAGCGAGGCCGACATCTGGACAACATGACCTGGATGCAAGTGGACCGCGAG
ATCAACAACACTACCGAGGACCATCTCCGCTGCTGGAGGACAGCCAGAACAGCAGGAGAACAGAGAACAGGACCTGCT
GGAGCTGGACAAGTGGAAACAACCTGTGGAACTGGTTCGACATCGCAACTGGCTGTGGTACATCAAGATCTCATCATGA
TCGTGGCGCCCTGATCGGCCCTGCGCATCATCTCGCCGTGCTGAGCATCGTAACCGCGTGCAGGCCAGGGCTACAGCCCC
CTGAGCTTCCAGACCCCTGACCCCAAGCCCCCGGCCCTGGACCGCTGGCGGATCGAGGAGGAGGGCGGAGCAGGA
CCCGCAGCGCAGCATCCGCTGGTGAAGCGGCTTCTGAGGCTGGCTGGACGACTGCGCAGGCTGTGCTTCTAGCT
ACCACCGCCCTGCGCAGCTTCACTGATCGCCGTGCGCCGCTGGAGCTGCTGGGCCACAGCAGCCCTGCCGGCTGCGAG
CGCGGCTGGAGATCCTGAAGTACTGGCAGGCCGTGGTGCAGTACTGGGCCCTGGAGCTGAAGAACAGGCCATCAGCCC
CCTGGACACCCTGCCATCGCCATCGCGTGGCGAGGGCACCGACCGCATCATCGAGCTGGTGCAGCGCATCTGCCCGCCATCC
TGAACATCCCCCGCCGATCCGCCAGGGCTTGTGGCGAGGGCTGCTGTAA

FIGURE 20

>Envgp160_TV1_C_ZAw (SEQ ID NO:50)

ATGAGAGTGATGGGGACACAGAAGAATTGTCAACAATGGTGGATAATGGGCATCTTAGGCTTCTGGATGCTAATGATTG
TAACACGGAGGACTTGTGGGTACAGTCACTATGGGTACCTGTGTGGAGAGAAGCAAAACACTCTATTCTGTGCAT
CAGATGCTAACAGCATATGAGACAGAAGTCATAATGTCTGGCTACACATGCTTGTGTACCCACAGACCCCACCCACAA
GAAATAGTTTGGGAAATGTAACAGAAAATTAAATATGTGAAATAACATGGCAGATCAGATGCATGAGGATATAAT
CAGTTATGGGATCAAAGCCTAACGCCATGTGAAAGTGCACCCACTGTGTCACTTAAACTGTACAGATAACAAATG
TTACAGGTAATAGAACTGTTACAGGTAATACAAATGATAACATATTGCAATGCTACATATAAGTATGAAGAAATGAAA
AATTGCTCTTCAATGCAACCACAGAATTAAGAGATAAGAACATAAAGAGTATGCACTCTTTATAAACTGTATAGT
ACCACTTAATGAAAATAGTAACAACTTACATATAGATTAATAAATTGCAATACCTCAACCTAACACAAGCCTGTCCA
AGGTCTCTTTGACCCGATTCCCTACATTACTGTGTCCAGCTGATTATGCGATTCTAAAGTGTAAATAAAAGACATTC
AATGGGACAGGACCATGTTATAATGTCAGCACAGTACAATGTCACATGGAATTAAAGCAGTGGTATCAACTCAACTACT
GTAAATGGTAGTCTAGCAGAAGGATAATAATTAGATCTGAAAAATTGACAGAGATAACCAAAACAATAATAGTAC
ATCTTAATGAATCTGTAGAGATTAATTGTACAAGGCCAACATAATAACAGGAAAAGTGTAAAGGATAGGACCAAGGACAA
GCATTCTATGCAACAAATGACGTAATAGGAAACATAAGACAGCATTGTAACTTGTAAATAAAACATAAAATTGAA
TTTACAACAGGTAATGAAAAATTAGGAGAGCATTTCCCTAAATAAAACATAAAATTGAAACACATGCAGGAGGGGATC
TAGAAATTACAATGCACTAGTTAATTGTAGAGGAGAATTCTATTGCAATACATCAAACCTGTTAATAGTACATAC
TACCTTAAGAATGGTACATACAAATAACATGGTAATTCAAGCTTACCCATCACACTCCAATGCAAAATAAAACAAATTG
ACGCATGTGCAAGGGTAGGACAAGCAATGTGATGCCCTCCATTGAGGAAACATAACATGTAGATCAAACATCACAG
GAATACTATTGACACGTGATGGGGATTAAACACAAACACGACACAGAGGAGACATTGACACCTGGAGGGAGGAG
ATGAGGGATAACTGGAGAATGAAATTATAAAATATAAAAGTGGTAGAAATTAGCATTGGGAAATAGCACCACAAAGC
AAAAAGAAGAGTGGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
GCACTATGGCGCAGGCTCAATAACGCTGACGGTACAGGCCAGACAACCTGTTGTCTGGTATAGTGCACAGCAAAGCAAT
TTGCTGAAGGCTATAGAGGCGCAACAGCATATGTCACACTCACAGTCTGGGCTTAAAGCAGCTCCAGGGAGAGTCT
GGCTATAGAAAGATAACCTAAAGGATCAACAGCTCTAGGATTGGGCTGCTCTGAAGACTCATCTGCACCACTGCTG
TGCCTTGGAACTCCAGTTGGAGTAATAAAATCTGAAGCAGATATTGGATAACATGACTGGATGCACTGGGATAGAGAA
ATTAATAATTACACAGAAACAAATTCTAGGTTGCTTGAAGACTCGAAAACAGCAGGAGGAGAATGAAAAAGATTATT
AGAATTGGACAAGTGGATAATCTGGAATTGGTTGACATATCAAACACTGGCTGTGGTATATAAAATATTCAAATGA
TAGTAGGAGGCTTGATAGTTAAAGATAATTGGCTGTCTCTATAGTGAATAGAGTTAGGCAGGGATACTCACCT
TGTGTCATTTCAGACCCCTAACCCAGCCGAGGGACTCGACAGGCTGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGA
CAGAGACAGATCCATACGATTGGTAGCGGATTCTGTGCTTGCCTGGGACGATCTGCGGAGCCTGTGCCTCTCAGCT
ACCACCGCTTGAGAGACTCATATTAAATGCACTGAGGGCAGTGGAACTCTGGGACACAGCAGTCTCAGGGACTACAG
AGGGGGTGGGAGATCCTAAAGTATCTGGGAAGTCTGTGCACTATTGGGCTTAGAGCTAAAAAGAGTGTATTAGTCC
GCTTGATACCATAGCAATAGCAGTAGCTGAAGGAACAGATAAGGATTATAGAATTGGTACAAAGAATTGTAGAGCTATCC
TCAACATACCTAGGAGAATAAGACAGGGCTTGAAGCAGCTTGCTATAA

FIGURE 21

>Gag_TV1_C_ZAopt (SEQ ID NO:51)

ATGGGCGCCCGGCCAGCATCCTGAGCGGGCAAGCTGGACAAGTGGGAGCGCATCCGCCTGCGCCCCGGCGCAAGAA
GCACATACATGCTGAAGCACCTGGTGTGGCCAGCCGAGCTGGAGCGCTTCGCCCTGAACCCCCGCTGCTGGAGACCA
GCGAGGGCTGCAAGCAGATCATCAAGCAGCTGGCAGCCCCCTGCAGACCGCACCAGGAGGAGCTGCGCAGCCGTGTTAAC
ACCGTGGCACCCCTGTAECTGCGTGCACAAGGGCATCGAGGTGCGCAGACCCAAGGAGGCCCTGGACAAGATCGAGGAGGA
GCAGAACAAAGTGCCAGCAGAACGGCCAGCAGGCCAACGGCCGAGAGAAGGTGAGGCCAGAACTACCCCATCGTGCAGA
ACGCCAGGGCAGATGGTGCACCAGGCATCAGCCCCCGCACCCCTGAACGCCCTGGATCAAGGTGATCGAGGAGAACGCC
TTCAACCCCGAGGAGATCCCCATGTTCAACGGCCCTGAGCGAGGGGCCACCCCCCAGGACCTGAACACCATGCTGAACAC
CGTGGGCGGCCACCAGGCCCATGCAAGATGCTGAAGGACACCATCAACGAGGAGGCCGAGTGGGACCGCACCCACC
CCGTGCACGCCGGCCCCGTGGCCAGATGCGCAGGCCCGCGGAGCAGATCGCCGCCACCACCGCACCCCTG
CAGGAGCAGATGCCCTGGATGACCAGAACCCCCCATCCCCGTGGAGGACATCTACAAGCGCTGGATCATCTGGCCT
GAACAAGATGTCGCGCATGTAACGCCCGTAGGCATCTGGACATCAAGCAGGGCCCCAAGGAGCCCTCCGCGACTACG
TGGACCGCTTCAAGACCCCTGCGCCGAGCAGGCCACCCAGGACGTAAGAAACTGGATGACCGACACCCCTGCTGGTG
CAGAACGCCAACCCCGACTGCAAGACCATCCTGCGCCCTGGCCCCGGCCAGCCTGGAGGAGATGATGACCGCCTG
CCAGGGCGTGGCGCCCCAGCCACAAGGCCCGCTGCTGGCCAGGCCATGAGCCAGGCCAACAGCAACATCCTGGTGC
AGGCAGCAAATTCAAGGGCAGCAACCGCATCATCAAGTGTCTCAACTGCGCAAGGTGGCCACATCGCCCGCACTGC
CGGCCCTGGCAAGATCTGGCCCAGCACAAGGGCCGGCAACTTCCCTGCAGAACCGCCCCGAGGCCACCGCCCCC
CCGGCAGGCCACCGCCCCCGCCAGAGCTCCGCTTCGAGGAGACCACCCCGTGGCCAGGCCAACAGGAGAGCGC
GAGCCCCCTGACCAGCCTGAAGAGCCTGTTGGCAGCGACCCCTGAGGCCAGTAA

FIGURE 22

>Gag_TV1_C_Zawt (SEQ ID NO:52)

ATGGGTGCGAGAGCGTCAATATTAAGCGGCGGAAAATTAGATAAAATGGGAAGAATTAGGTTAACGCCAGGGGGAAAGAA
ACATTATATGTAAAACATCTAGTATGGGCAAGCAGGGAGCTGGAAAGATTGCACTTAACCTGCCCTGTTAGAAACAT
CAGAAGGCTGAAACAAATAATAAAACAGCTACAACCAGCTTCAGACAGGAACAGAGGAACCTTAGATCATTATTCAAC
ACAGTAGCAACTCTATTGTGTACATAAAGGGATAGAGGTACGAGACCCAAGGAAGCCCTAGACAAGATAGAGGAAGA
ACAAAAACAATGTCAGCAAAAGCACACAGGCAAAAGCAGCTGACGAAAAGGTCACTAAATTATCCTATAGTACAGA
ATGCCCAAGGCAAATGGTACACCAAGCTATACACCTAGAACATTGAATGCATGGATAAAAGTAATAGAGGAAAGGCT
TTCAATCCAGAGGAAATACCCATGTTACAGCATTATCAGAAGGAGCCACCCCACAAGATTAAACACAATGTTAAATAC
AGTGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGATACCATCAATGAGGGCTGCAGAATGGGATAGGACACATC
CAGTACATGCAGGGCCTGTCACCAGGAGATGAGAGAACCAAGGGGAAAGTGACATAGCAGGAACACTACTAGTACCCCTT
CAGGAACAAATAGCATGGATGACAAGTAATCCACCTATTCCAGTAGAACAGATCTATAAAAGATGGATAATTCTGGTT
AAATAAAATAGTAAGAATGTATAGCCCTGTTAGCATTGGACATAAAACAAAGGGCAAAGAACCCCTTAGAGACTATG
TAGACCGGTTCTTAAACCTTAAGAGCTGAACAAGCTACACAAGATGTAAGAACAGCTACAGGCTGAGGCAATGAGGCAAG
CAAATGCGAACCCAGATTGTAAGACCATTAAAGAGCATTAGGACCAGGGCCTCATTAGAACAGATGACAGCATG
TCAGGGAGTGGAGGACCTAGCCATAAAGCAAGAGTGTGGCTGAGGCAATGAGGCAACAGAACAGTAACATACTAGTGC
AGAGAACAAATTAAAGGCTTAACAGAATTATAATGTTCAACTGTGGCAAAGTAGGGCACATAGCCAGAAATTGC
AGGGCCCTAGGAAAAGGCTGTTGAAATGTGGACAGGAAGGACACCAATGAAAGACTGTACTGAGAGGCAAG
TTTTTAGGGAAAATTGGCTTCCACAAGGGGAGGCCAGGAATTCCAGAACAGACCAGGCCAACAGCCCCAC
CAGCAGAACCAACAGCCCCACCAGCAGAGAGCTTCAGGTTGAGGAGACAACCCCGTCCGAGGAAGGAGAAAGAGAGG
GAACCTTTAACCTCCCTCAAATCACTTTGGCAGCGACCCCTGTCTCAATAA

FIGURE 23

>Gag_TV1_ZA_MRopt (SEQ ID NO:53)

GACATCAAGCAGGGCCCCAAGGAGCCTTCCGCGACTACGTGGACCGCTTCAAGACC

FIGURE 24

>Gag_TV1_ZA_MHRwt (SEQ ID NO:54)

GACATAAAACAAGGGCAAAAGAACCCTTTAGAGACTATGTAGACCGGTTCTTAAACC

FIGURE 25

>Nef_TV1_C_ZAopt (SEQ ID NO:55)

ATGGGCGGCAAGTGGAGCAAGCGCAGCATCGTGGCTGGCCGCGTGCAGCGCATGCCGCACCGAGCCGCCGC
CGAGGGCGTGGCGCCAGCCAGGACCTGGACGCCACGGGCCCTGACCAGCAGCAACACCCCCGCCACCAACGAGG
CCTGCGCCTGGCTGCAGGCCAGGAGGAAGGGCGACGTGGCTTCCCCTGCGCCCCCAGGTGCCCCCTGCGCCCCATG
ACCTACAAGAGCGCGTGGACCTGAGCTTCCTGAAGGAGAAGGGGGCTGGAGGGCTGATCTACAGCCGCAAGCG
CCAGGAGATCTGGACCTGTGGTGTACAACACCCAGGGCTTCTTCCCCTGACTGGCAGAACTACACCAGGGCCCCGGCG
TGCGCTTCCCCCTGACCTTCGGCTGTTCAAGCTGGTGGCCGTGGACCCCCGGAGGTGAAGGAGGCCAACGAGGGC
GAGGACAACCTGCCTGCTGCACCCATGAGCCAGCACGGGCCAGGACGAGGAGGGCTGAAGTGGAAAGTTCGA
CAGCCTGCTGGCCCACCGCCACATGGCCCGAGCTGCACCCGAGTACTACAAGGACTGCTGA

FIGURE 26

>Nef_TV1_C_Zawt (SEQ ID NO:56)

ATGGGAGGCAAGTGGTCAAAACGCAAGCATAGTTGGATGGCCTGCAGTAAGAGAAAAGAATGAGAAGAACTGAGCCAGCAGC
AGAGGGAGTAGGAGCAGCGTCTCAAGACTTAGATAGACATGGGCACCTACAAGCAGCAACACACCTGCTACTAATGAAG
CTTGTGCCTGGCTGCAAGCAACAAGAGGGAGGACGGAGATGTAGGCTTTCCACTCAGACCTCAGGTACCTTAAGACCAATG
ACTTATAAGAGTGCAAGTAGATCTCAGCTTCTTTTAAAGAAAAGGGGGACTGGAAGGGTTAATTACTCTAGGAAAAG
GCAAGAAAATCCTGATTTGTGGGTCTATAACACACAAGGCTCTCCCTGATGGCAAACACTACACATGGGGCCAGGGG
TCCGATTCCCAC TGACCTTTGGATGGTCTCAAGCTAGTACCA GGTGACCCAAGGGAGGTGAAAGAGGGCAATGAAGGA
GAAGACAAC TGTTGCTACACCTATGAGCCAACATGGAGCAGAGGATGAAGATAGAGAAGTATTAAAGTGGAAAGTTGA
CAGCCTCTAGCACACAGACACATGGCCCGAGCTACATCCGGAGTATTACAAAGACTGCTGA

FIGURE 27

>NefD125G_TV1_C_ZAopt (SEQ ID NO:57)

ATGGGGCGGAAGTGGAGCAAGCGCAGCATCGTGGGCTGGCCGCCGTGCGCGAGCGCATGCGCCGCACCGAGGCCGCCGC
CGAGGGCGTGGCGCCAGCCAGGACCTGGACCGCCACGGCGCCCTGACCAGCACAAACACCCCCGCCACCAACGAGG
CCTGCGCCTGGCTGCAGGCCAGGAGGACGGCGACGTGGGCTTCCCCGTGCGCCCCCAGGTGCCCCCTGCGCCCCATG
ACCTACAAGAGCGCCGTGGACCTGAGCTTCCTGAAGGAGAAAGGGCGGCCCTGGAGGGCCTGATCTACAGCCGCAAGCG
CCAGGAGATCTGGACCTGTGGGTGTACAACACCCAGGGCTTCTCCCCGGCTGGCAGAACTACACCAGCGGGCCCCGGCG
TGCCTTCCCCCTGACCTTCGGCTGGTCTTCAAGCTGGTGCCGTGGACCCCCCGCGAGGTGAAGGAGGCCAACGAGGGC
GAGGACAACGTGCTGCTGCACCCCAGGCCACGGCGCCAGGACGAGGACCGCGAGGTGCTGAAGTGGAAAGTCGA
CAGCCTGCTGGCCCACCGCCACATGGCCCGAGCTGCACCCCAGTACTACAAGGACTGCTGA

FIGURE 28

>p15RNaseH_TV1_C_ZAopt (SEQ ID NO:58)

ACCTTCTACGTGGACGGGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCTACGTGACCGACCGCGGCCAGAA
GATCGTGACCCCTGACCAACACCAACCAACCAAGAACGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGCAGCG
AGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGCATCATCAGCCCCAGCCCCACAAGAGCGACAGCGAGATCTTC
AACCAAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCCCACAAGGGCATCGGCCGCAA
CGAGCAGGTGGACAAGCTGGTGAGCAAGGGCATC

FIGURE 29

>p15RNaseH_TV1_C_Zawt (SEQ ID NO:59)

ACTTTCTATGTAGATGGAGCAACTAATAGGAAGCTAAAATAGGAAAAGCAGGGTATGTTACTGACAGAGGAAGGCAGAA
AATTGTTACTCTAACATAACACAACAAATCAGAAGACTGAGTTACAAGCAATTCAAGCTAGCTCTGCAGGATTCAAGGATCAG
AAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAGCACAACCAAGATAAGAGTGACTCAGAGATATT
AACCAAATAATAGAACAGTTAATAAACAAAGGAAGAATCTACCTGTACGGTACCGAGCACATAAGGAATTGGGGAAA
TGAACAAGTAGATAAAATTAGTAAGTAAGGGAATT

FIGURE 30

>p31Int_TV1_C_Zaopt (SEQ ID NO:60)

CGCAAGGTGCTGTTCTGGACGGCATCGACAAGGCCAGGAGGAGCACGAGCGCTACCACAGCAACTGGCGGCCATGGC
CAACGAGTTCAACCTGCCCGCATCGTGGCCAAGGAGATCGTGGCCAGCTGGCACAAGTGCAGCTGAAGGGCGAGGCCA
TCCACGGCCAGGTGGACTGCAGCCCCGGCATCTGGCAGCTGGACTGCACCCACCTGGAGGGCAAGATCATCCTGGTGGCC
GTGCACGTGGCCAGCGGCTACATGGAGGCCAGGTGATCCCCGGCAGAACCGGCGACTTCACCCAGCAGCCGTGAAGGCCGCT
GCTGGCCGGCCGCTGGCCCGTGAAGGTGATCCACACCGACAACGGCAGCAACTTCACCCAGCAGCCGTGGAGAGCATGAACAAG
GCTGGTGGGCCGGCATCCAGCAGGAGTTGGCATCCCCTACAACCCCCAGAGCCAGGGCGTGGTGGAGAGCATGAACAAG
GAGCTGAAGAAGATCATCGGCCAGGTGGCGACCAGGCCAGCACCTGAAGACGCCGTGCAGATGCCGTGTTCATCCA
CAACTTCAAGCGCAAGGGCGCATCGGGGCTACAGCGCCGGCAGCGCATCATGACATCATGCCACCGACATCCAGA
CCAAGGAGCTGAGAAGCAGATCATCCGATCCAGAACTTCCCGTGTACTACCGCGACAGCCCGAACCCCATCTGGAAG
GGCCCGCCGAGCTGCTGTGGAAGGGCGAGGGCGTGGTGGTGTACGAGGACAAGGGGACATCAAGGTGGTGGCCAGGACGAGGAC
CAAGGCCAAGATCATCCGCACTACGGCAAGCAGATGGCCGGCGCGACTGCGTGGCCGGCGCAGGACGAGGAC

FIGURE 31

>p31Int_TV1_C_Zawt (SEQ ID NO:61)

AGGAAAGTGTGTTCTAGATGGAATAGATAAAGCTCAAGAAGAGCATGAAAGGTACCACAGCAATTGGAGAGCAATGGC
TAATGAGTTAACATCTGCCACCCATAGTAGCAAAAGAAATAGTAGCTAGCTGTGATAATGTCAGCTAAAAGGGGAAGCCA
TACATGGACAAGTCGACTGTAGTCAGGGATATGCCAATTAGATTGTACCCATTAGAGGGAAAATCATCCCTGGTAGCA
GTCCCATGTAGCTAGTGGCTACATGGAAGCAGAGGTATCCCAGCAGAAACAGGACAAGAACAGCATATTATTTA
ATTAGCAGGAAGATGCCAGTCAAAGTAATACATACAGACAATGGCAGTAATTTCACCAGTACTGCAGTTAAGGCAGCCT
GTTGGTGGGCAGGTATCCAACAGGAATTGGAATTCCCTACAATCCCCAAAGTCAGGGAGTGTTAGAATCCATGAATA
GAATTAAAGAAAATAATAGGACAAGTAAGAGATCAAGCTGAGCACCTTAAGACAGCAGTACAATGGCAGTATTCA
CAATTAAAAGAAAAGGGGAATTGGGGGTACAGTGCAGGGAAAAGAATAATAGACATAATAGCAACAGACATACAAA
CTAAAGAATTACAAAAAAATTATAAGAATTCAAATTCGGGTTATTACAGAGACAGCAGAGACCCATTGGAAA
GGACCAGCCGAACTACTCTGGAAAGGTGAAGGGTAGTAGTAATAGAAGATAAAGGTGACATAAAGGTAGTACCAAGGAG
GAAAGCAAAATCATTAGAGATTATGGAAAACAGATGGCAGGTGCTGATTGTGGCAGGTGACAGGATGAAGAT

FIGURE 32

>Pol_TV1_C_ZAopt (SEQ ID NO:62)

TTCTTCGCGAGAACCTGGCCTCCCCAGGGCGAGGCCCGAGTCCCCCCCAGCAGACCCCGGCCAACAGCCCCAC
CAGCCGCACCAACAGCCCCACCAGCGCAGCTGCAGGTGCGCGCGACAACCCCCCGCGAGGAGGGCGAGCGAGG
GCACCTTCAACTCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCC
CTGCTGGACACCGGCCGACGACACCGTGTGGAGGAGATCGACCTGCCCGCAAGTGGAAAGGCCAAGATGATCGGCC
CATCGGCCGCTTCATCAAGGTGCGCAGTACGACCAGATCTGATCGAGATCTGGCGCAAGAAGGCCATCGGCCACCGTGC
TGGTGGGCCACCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGTGCAACCTGAACCTCCCCATCAGC
CCCATCGAGACCGTGCCTGAAGCTGAAGCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGGCCCTGACCGAGGAGAA
GATCAAGGCCCTGACCCGCATCTGCAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCGACAACCCCTACA
ACACCCCCGTGTTGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGC
ACCCAGGACTCTGGAGGTGCACTGGCATCCCCACCCCGCCGGGCTGAAGAAGAAGAAGAGCGTGAACCGTGC
CGTGGCGACGCCCTACCTCAGCGTCCCCCTGGACGAGAGCTTCCGAAGTACACCCCTTCACCATCCCCAGCATCAACA
ACGAGACCCCGGATCCGCTACCAAGTACACCGTGTGCCAGGGCTGGAGGGCAGGCCATCTTCAGAGCAGC
ATGACCAAGATCTGGAGGCCCTCGCGCCAAGAACCCGACATCGTACATCCAGTACATGGACGACCTGTACGTGG
CAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGGCGAGGACACTGCTGAAGTGGGCTTCACCA
CCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGTGGATGGCTACAGGCTGCCACCCGACAAGTGGACCGTGCAGCCC
ATCCTGGTGCCCGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGGAGGCCAGAT
CTACCCCGGCATCAAGTGCAGCTGTGCAAGCTGCTGCCGCGCCAAGGCCCTGACCGACATCGTGGGCCCTGACCG
AGGAGGCCAGCTGGAGCTGGCGAGAACCCGAGATCTTGCGGAGGCCCTGTCACGGCGTGTACTACGACCCAGCAAG
GACCTGATCGCCGAGATCCAGAACAGGGCCAGGAGCTGGACCTACCAAGATCTACCAAGGAGCCCCCTTAAGAACCTGAA
GACCGCAAGTACGCCAAGATGCCAGGCCACCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAAGAAGATGCCA
TGGAGAGCATCGTACGGGCAAGACCCCCAAGTCTGGCTGCCATCCAGAAGGAGACCTGGAGACCTGGTGGACC
GACTACTGGCAGGCCACCTGGATCCCCAGTGGAGATTGGTGAACACCCCCCTGGTGAAGCTGTGGTACAGCTGGA
GAAGGACCCCCTGCGCGTGGAGACCTTCACTGTGGACGGCGCCACCAACCGCGAGGCCAGATCGGAAGGCCGGT
ACGTGACCGACCGCGGCCAGAACGATCGTACCCCTGACCAACACCACCAAGAACGCCAGCTGCAAGGCCATCCAG
CTGGGCCCTGCGAGCACGGCAGCGAGGTGACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCC
CGACAAGAGCGACAGCGAGATCTCAACCGACATCGACAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGTGC
CCGCCACAGGGCATCGCGGCCACGAGCAGGAGCTGGACAAGCTGGTGAAGGCCAGGTGCTGGGCCCTGGAC
GGCATCGACAAGGCCAGGAGGAGCACGGCTACCAACAGCAACTGGCGCCATGGCAACGAGTTCAACCTGCC
CATCGTGGCCAAGGAGATCGTGGCCAGCTGCCACAGTGGCAGCTGAAGGGCGAGGCCATCCACGGCCAGGTGGACTGCA
GCCCGGCATCTGGCAGCTGGACTGCACCCACCTGGAGGGCAAGATCATCTGGTGGCCGTGACGTGGCAGCGGCTAC
ATGGAGGCCAGGGTACCCCCCGAGACCGCCAGGAGACCGCTACTTCATCTGAAGCTGGCGGCCCTGGCC
GAAGGTGATCCACACCGACAACGGCAACTTACCAAGCAGGCCGTGAAGGCCCTGCTGGTGGGCCCTGGCATCCAGC
AGGAGTTGGCATCCCCCTACAACCCAGGCCAGGGCGTGGAGAGCATGAACAAGGAGCTGAAGAAGATCATGGC
CAGGTGCGGCCAGGCCAGGCCAGCTGAAGACCGCCGTGCAAGATGGCGTGTTCATCCACAACCTCAAGGCCAAGGGCG
CATCGCGGCCACAGGCCGCCAGCGATCATCGACATCATGCCACCGACATCCAGACCAAGGAGCTGCAAGCAGA
TCATCCGACATCCAGAACCTCCGCGTGTACTACCGCGACAGCCCGACCCATCTGGAGGGCCCGAGCTGCTGTGG
AAGGGCGAGGGCGTGGTGGTGAATCGAGGACAAGGGCGACATCAAGGTGGTGGCCCGCAAGGCCAAGATCATCCGCGA
CTACGGCAAGCAGATGGCCGGCGCCGACTGGTGGCCGGCGAGGACGAGGAC

FIGURE 33

>Pol_TV1_C_Zawt (SEQ ID NO:63)

TTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGAGGCCAGGGAAATTCCCTCCAGAACAGACCAGAGCCAACAGCCCCAC
CAGCAGAACCAACAGCCCCACCAGCAGAGCTTCAGGTTCGAGGAGACAACCCCGTGCAGGAGAAAGGAGAAAGAGAGG
GAACCTTTAACCTCCCTCAAATCACTCTTGGCAGCGACCCCTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCT
CTCTTAGACACAGGAGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGGAACCAAAATGATAGGGGG
AATTGGAGGTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTGTGGAAAAAAAGGCTATAGGTACAGTAT
TAGTAGGGCTACACCACTGCAACATAATTGGAAGAAACTGTTAATCTCAGCTGGATGCACACTAAATTTCCTAATTAGT
CCTATTGAAACTGTACCACTGAAATTAAACCAAGGAATGGATGGCCAAAGGTCAAACAATGGCCATTGACAGAAAGAAA
AATAAAAGCATTAACAGCAATTGTGAGGAAATGGAGAAGGAAGGAAAATTACAAAAATTGGGCTGATAATCCATATA
ACACTCCAGTATTGCCATAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAGGGAACTCAATAAAAGA
ACTCAAGACTTTGGGAAGTTCATTAGGAATACCACCCCAGCAGGATTAAGGAAAATCAGTGACAGTGTAGA
TGTGGGGATGCATATTTCAGTTCTTAGATGAAAGCTTCAGGGAAATATCTGCAATTCCATACCTAGTATAAACAA
ATGAAACACCAGGGATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAGGATCACCAGCAATATTCCAGAGTAGC
ATGACAAAAATCTTAGGCCCTCAGAGCAAAATCCAGACATAGTTATCTAATATGGATGACTTGTATGTAGG
ATCTGACTTAGAAATAGGGCAACATAGAGCAAAATAGAAGAGTTAAGGAAACATTATTGAAATGGGATTACACAC
CAGACAAGAAACATCAAAAGAACCCCCATTCTTGGATGGGATGAACTCCATCTGCAAATGGACAGTACACCT
ATACTGCTGCCAGAAAGGATAGTTGACTGTCAATGATATACAGAAGTTAGTGGAAAATTAAACTGGCAAGTCAGAT
TTACCCAGGGATTAAAGTAAGGCAACTCTGTAACCTCTCAGGGGGGCAAAAGCATAACAGACATAGTACCAACTG
AAGAAGCAGAATTAGAATTGGCAGAGAACAGGGAAATTAAAGAGAACAGTACATGGAGTATAATTGATCCATCAAAA
GACTTGATAGCTGAAATACAGAACAGGGCATGAACAATGGACATATCAAAATTATCAAGAACATTAAATCTGAA
AACAGGGAAGTATGCAAAATGAGGACTACCCACACTAATGATGAAACAGTTAACAGAGGCACTGCAAAAATAGCCA
TGGAAAGCATAGTAATATGGGAAAGACTCTAAATTAGACTACCCATCAGAACATGGAGACATGGTGGACA
GACTATTGGCAAGCCACCTGGATCCCCTGAGTGGAGTTGTTAATACCCCTCCCTAGTAAATTATGGTACCAACTAGA
AAAAGATCCCATAGCAGGAGTAGAAACTTCTATGTAGATGGAGCAACTAATAGGAAAGCTAAATAGGAAAGCAGGGT
ATGTTACTGACAGAGGAAGGCCAGAAAATTGTTACTCTAACTAACACAACATCAGAACAGTGTGAGTTACAAGCAATTG
CTAGCTCTGCAGGATTCAAGGATCAGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCTCAAGCACAACC
AGATAAGAGTCACTCAGAGATATTAAACAAATAATAGAACAGTTAACAAACAAGGAAAGAATCTACCTGTATGGTAC
CAGCACATAAGGAATTGGGAAATGAAACAGTAGATAAAATTAGTAAGTAAAGGAAATTAGGAAAGTAAATGGTGTGTTCTAGAT
GGAATAGATAAGCTCAAGAACAGCATGAAAGGTACACAGCAATTGGAGAGCAATGGCTAATGAGTTAACCTGCCACC
CATAGTAGCAAAAGAAATAGTAGCTAGCTGTGATAATGTCAGCTAAAGGGGAAGCCATACATGGACAAGTCGACTGTA
GTCCAGGGATATGGCAATTAGATTGTACCCATTAGAGGAAAATCATCTGGTAGCAGTCCATGTAGCTAGTGGCTAC
ATGGAAGCAGAGGTTATCCCAGCAGAACAGGACAAGAACAGCATATTATATTAAATTAGCAGGAAGATGCCAGT
CAAAGTAATACATACAGACAATGGCAGTAATTTCAGTACTGCAAGTTAACGGCAGCTGGTGGCAGGTATCCAAC
AGGAATTGGAAATTCCCTACAATCCCCAAAGTCAGGGAGTGGTAGAATCCATGAATAAAAGAATTAAAGAAAATAATAGGA
CAAGTAAGAGATCAAGCTGAGCACCTTAAGACAGCAGTACAAATGGCAGTATTCTACATCAAAATTAAAGAAAAGGGGG
AATTGGGGGTACAGTCAGGGAAAGATAATAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAACAAA
TTATAAGAATTCAAATTTCGGGTTATTACAGAGACAGCAGAGACCCCTATTGGAAAGGACCAGCCGAACACTCTGG
AAAGGTGAAGGGTAGTAGTAAAGAGATAAAGGTGACATAAAGGTAGTACCAAGGAGGAAAGCAAAATCATTAGAGA
TTATGGAAAACAGATGGCAGGTGCTGATTGTGGCAGGTGGACAGGATGAAGAT

FIGURE 34

>Prot_TV1_C_ZAopt (SEQ ID NO:64)

CCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCCCTGCTGGACACCGG
CGCCGACGACACCGTGTGGAGGAGATCGACCTGCCCGCAAGTGGAAAGCCAAGATGATCGGCGGCATCGGCGGCTTCA
TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCCTGAACCTTC

FIGURE 35

>Prot_TV1_C_ZAwt (SEQ ID NO:65)

CCTCAAATCACTCTTGGCAGCGACCCCTTGTCATAAAAAGTAGAGGCCAGATAAAGGAGGCCTCTTAGACACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTGCCAGGGAAATGGAAACCAAAATGATAGGGGAATTGGAGGTTTA
TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGGCCTACA
CCAGTCAACATAATTGGAAGAAATCTGTTAACTCAGCTTGGATGCACACTAAATT

FIGURE 36

>Protina_TV1_C_ZAopt (SEQ ID NO:66)

CCCCAGATCACCCGTGGCAGCGCCCCCTGGTGACCATCAAGGTGGAGGGCCAGATCAAGGAGGCCCTGCTGGCCACCGG
CGCCGACGACACCGTGCTGGAGGAGATCGACCTGCCCGCAAGTGGAAAGCCAAGATGATCGCGGCATCGGCCTTCA
TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGGCAAGAAGGCCATCGGCACCGTGTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCCTGAACCTTC

FIGURE 37

>Protina_TV1_C_Zawt (SEQ ID NO:67)

CCTCAAATCACTCTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCTCTTTAGGCCACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTGCCAGGGAAATGGAAACCAAAATGATAGGGGAATTGGAGGTTTA
TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGGCCTACA
CCAGTCACATAATTGGAAGAAATCTGTTAACTCAGCTTGGATGCACACTAAATT

FIGURE 38

>ProtinaRTmut_TV1_C_ZAopt (SEQ ID NO:68)

CCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCCCTGCTGGCCACCGG
CGCCGACGACCCGTGCTGGAGGAGATCGACCTGCCGGCAAGTGGAAAGCCAAGATGATCGGCGGCATCGGCGGCTTCA
TCAAGGTGCGCCAGTACGACCAAGATCCTGATCGAGATCTCGGCAAGAAGGCCATCGGACCGTGCTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGACCCCTGAACTTCCCCTCAGCCCCATCGAGACCGT
GCCCCGTGAAGCTGAAGGCCGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGA
CCGCCATCTCGGAGGAGATGGAGAAGGAGGGCAAGATACCAAGATCGGCCCGACAACCCCTACAACACCCCCGTGTT
GCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTG
GGAGGTGCACTGGGCATCCCCCACCCCGCCGCTGAAGAAGAAGAAGCGTGACCGTGCTGGACGTGGCGACGCCT
ACTTCAGCGTGCCTCTGGACAGAGACTTCGCAAGTACACCGCCTTCACCCATCCCCAGCATCAACAACGAGACCCCCGGC
ATCCGCTACCACTGACGACTGGCCCAAGGGCTGAGGGCAGCCCCGCCATCTCCAGAGCAGCATGAAGATCCT
GGAGCCCTTCGCGCCAAGAACCCCGACATCGTATCTACCGAGCCCCCTGTACGTGGCAGCGACCTGGAGATCGGCC
AGCACCGCGCCAAGATCGAGGGAGCTGGCGAGCACCTGCTGAAGTGGGGCTTCACCCACCCCGACAAGAAGCACCAGAAG
GAGCCCCCTTCCTGCCCCATCGAGCTGCAACCCGACAAGTGGACCGTGCAGCCCATCTGCTGCCAGAGACAGCTG
GACCGTGAACGACATCCAGAAGCTGGGGCAAGCTGAACTGGCCAGCCAGATCTACCCCGGCACTCAAGGTGCGCCAGC
TGTGCAAGCTGCTGCGCGGCCAACGGCCATCGACCGACATCGTCCCCCTGACCGAGGGAGGGCAGCTGGAGCTGGCCAG
AACCGCGAGATCTGCGCGAGCCCCCTGACCGCGTGTACTACGACCCAGCAAGGACCTGATGCCAGATCCAGAAGCA
GGGCCACGAGCAGTGGACCTTACCAAGATCTACCAAGGAGCCCCCTCAAGAACCTGAAAGACCCGCAAGTACGCCAAGATGCGCA
CCACCCACACCAACGACGTGAAGCAGCTGACCGAGGGCGTGCAAGAGATGCCATGGAGAGCATCGTATCTGGGGCAAG
ACCCCCAAGITCCGCTGCCCATCCAGAAGGGAGACCTGGGAGACCTGGTGACCGACTACTGGCAGGCCACCTGGATCCC
CGAGTGGGAGTTCTGTAACACCCCCCCCCTGGTGAAGCTGGTACCGAGCTGGAGAAGGACCCCATGCCGGCGTGGAGA
CCTTCTACGGGACGGGCCACCAACCGCAGGCCAGATGCCAAGGCGCTACGTGACCGAGCCGGCCAGAG
ATCGTGACCCCTGACCAACACCACCAACCAAGGCCAGCTGACGGCCATCCAGCTGGCCCTGCAAGGACAGCGCAGCGA
GGTGAACATCGTGACCGACAGCCAGTACGCCCTGGCATCTCAGGCCAGGCCACAAGAGCGACAGCGAGATCTTCA
ACCAAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCGCCACAAGGGCATCGGGCAAC
GAGCAGGTGGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG

FIGURE 39

>ProtinaRTmut_TV1_C_ZAwT (SEQ ID NO:69)

CCTCAAATCAGCTTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCTCTCTAGGCCACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGAAACCAAAATGATAGGGGAATTGGAGGTTTA
TCAAAGTAAGAACAGTATGATCAAATACTTATAGAAATTGTGAAAAAGGCTATAAGTACAGTATTAGTAGGGCTACA
CCAGTCACATAATTGGAGAAAATCTGTTAAGTCACTAGCTGGATGCACACTAAATTTCACATTAGTCCTATTGAAACTGT
ACCAAGTAAACAGGAAATGGATGGCCAAAGGCTAAACAAATGCCATTGACAGAAGAAAAATAAAAGCATTAA
CAGCAATTGTGAGGAAATGGAGAAGGAAAGGAAATTACAAAAATTGGGCTGATAATACACTCCAGTATT
GCCATAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAAGGAACTCAATAAAAGAACTCAAGACTTTG
GGAAGTTCAATTAGGAATACCACACCAGCAGGATTAAAAAGAAAAATCAGTGACAGTGCTAGATGTGGGGATGCAT
ATTTCAGTCCTTATGAGTGAAGCTCAGGAAATACTGCATTACCTAGTATAAACAAATGAAACACCAGG
ATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACAGCAATATTCCAGAGTACGATGACAAAAATCTT
AGGCCCTCAGAGCAAAATCCAGACATAGTTATCTATCAAGCCCCTTGTATGTAGGATCTGACTTAGAAATAGGC
AACATAGAGCAAAATAGAAGAGTTAAGGGAACATTATTGAAATGGGATTACACACCAGACAAGAAACATCAAAA
GAACCCCCATTCTCCCATCGAACTCCATCCTGACAAATGGACAGTACAACCTATACTGCTGCCAGAAAGGATAGTG
GACTGTCAATGATATACAGAAGTTAGTGGAAATTAAACTGGCAAGTCAGATTACCCAGGGATTAAAGTAAGGCAAC
TCTGTAAACTCTCAGGGGGCAAAGCACTAACAGACATAGTACCAACTAATGAAAGCAGAATTAGAATTGGCAGAG
AACAGGGAAATTAAAGAGAACAGTACATGGAGTTATTTATGATCCATCAAAAGACTTGATAGCTGAAATACAGAAACA
GGGGCATGAACAATGGACATATCAAATTATCAAGAACCATTTAAACAGGGCACTGCAAAATAGCCATGAAAGCAGTGTAA
CTACCCACACTAATGATGTAACACAGTTAACAGAGGCAGTGCAAAATAGCCATGAAAGCAGTGTAAATGGGAAAG
ACTCCTAAATTAGACTACCCATCCAAAAAGAAACATGGGAGACATGGTGGACAGACTATTGGCAAGCCACTGGATCCC
TGAGTGGAGTTGTTAATACCCCTCCCTAGTAAATTATGGTACCAACTAGAAAAAGATCCCATAGCAGGAGTAGAA
CTTTCTATGTAGATGGAGCAACTAATAGGGAAAGCTAAAATAGGAAAGCAGGGTATGTTACTGACAGAGGAAGGAGAA
ATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAAGCAATTGCTCTGCAGGATTCAAGGATCAGA
AGTAAACATAGTAACAGACTCACAGTATGCTTACAGCACAACCAGATAAGAGTGACTCAGAGATATTAA
ACCAAAATAATAGAACACTTAATAACAAAGGAAAGATCTACCTGTCATGGTACCAAGCACATAAAGGAATTGGGGAAAT
GAACAAGTAGATAAATTAGTAAGTAAGGAAATTAGGAAAGTGTG

FIGURE 40

>ProtwtRTwt_TV1_C_ZAopt (SEQ ID NO:70)

CCCCAGATCACCCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCAGATCAAGGAGGCCCTGCTGGACACCGG
CGCCGACGACACCGTGTGGAGGAGATCGACCTGCCCGCAAGTGGAAAGGCCAAGATGATCGCGGCATCGCGGCTTC
TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGCAAGAAGGCCATCGCACCGTGCTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCCTGAACCTCCCCTCAGCCCATCGAGACCGT
GCCCGTGAAGCTGAAGCCCGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCTGACCAGGGAGAAGATCAAGGCCCTGA
CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCGACAACCCCTACAACACCCCCGTGTT
GCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTC
GGAGGTGAGCTGGGATCCCCCACCCCGCCGCTGAAGAAGAAGAGCGTGACCGTGCTGGACGCTGGGCGACGCCT
ACTTCAGCGTGCCTCTGGACGAGAGCTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGC
ATCCGCTACCACTGACAGCTGCTGCCCAAGGGCTGGAAAGGGCAGGCCGCACTTCCAGAGCAGCATGA^{CA}AGATCCT
GGAGCCCTTCCGCGCCAAGAACCCGACATCGTATCTACCACTGACATGGACGACCTGTACGTGGGAGCGACCTGGAGA
TCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGTGGGCTTACCCACCCCGACAAGAACGAC
CAGAAGGAGCCCCCTTCCCTGTGGATGGCTACGAGCTGACCCCGACAAGTGGACCGTGAGCCCATCTGCTGCCGA
GAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGGGCAAGCTGAACCTGGGCAAGGCCAGATCTACCCGGCATCA
AGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCCAAGGCCCTGACCGACATCGTCCCCCTGACCGAGGAGGCCAGCTG
GAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGGCCCTGACCGCGTGACTACGACCCAGCAAGGACCTGATCGCGA
GATCCAGAAGCAGGGCCACGAGCAGTGGACCTACCAAGATCTACCACTGACCGAGGCCCTTAAGAACCTGAAGACCGGCAAGTACG
CCAAGATGCGCACCAACCAACGACGCTGAAGCAGCTGACCGAGGCCCTGACAGAACATGGAGAGCATCGT
ATCTGGGCAAGACCCCAAGTCCGCTGCCATCCAGAAGGAGACCTGGAGACCTGGTGGACCCGACTACTGGCAGGC
CACCTGGATCCCGAGTGGAGTTCTGTAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAAGCTGGAGAACGGACCCATCG
CCGGCGTGGAGACCTCTACGTGGACGGCGCCAACCGCGAGGCCAAGATCGCAAGGCCGCTACGTGACCGACCG
GGCCGCCAGAAGATCGTGAACCTGACCAACACCAACCCAGAAGAACGGAGCTGAGGCCATCCAGCTGGCCCTGAGGA
CAGCGGCAGCGAGGTGAACATCGTACGCCAGTACGCCCTGGCATCCAGGCCAGCCGACAAGAGCGACA
GCGAGATCTCAACCAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCGCCACAAGGGC
ATCGCGGCAACGAGCAGGTGGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG

FIGURE 41

>ProtwtRTwt_TV1_C_Zawt (SEQ ID NO:71)

CCTCAAATCACTCTTGGCAGCGACCCCTGTCTCAATAAAAGTAGAGGCCAGATAAAGGAGGCTCTTAGACACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGGAAACAAAAATGATAGGGGGATTGGAGGTTTA
TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTGTGGAAAAAGGCTATAGGTACAGTATTAGTAGGGCCTACA
CCAGTCACATAATTGGAAGAAATCTGTTAACCTCAGCTGGATGCACACTAAATTTCAATTAGTCCTATTGAAACTGT
ACCAGTAAAATTAAAACCAGGAATGGATGGCCAAAGGTCAAACAAATGGCCATTGACAGAAGAAAAAATAAAGCATTAA
CAGCAATTGTGAGGAAATGGAGAAGGAAGGAAAATTACAAAATTGGGCTGATAATCCATATAACACTCCAGTATT
GCCATAAAAAGAAGGACAGTACTAAGTGAGAAAATTAGTAGATTTCAGGAACTCAATAAAAGAACTCAAGACCTTGG
GGAAGTCAATTAGGAATACCAACCCCAGCAGGATTAAAAAGAAAAAATCAGTGACAGTGTAGATGTGGGGATGCAT
ATTTTCAGTTCTTTAGATGAAAGCTTCAGGAAATATACTGCATTCACTACCTAGTATAACAAATGAAACACCCAGGG
ATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAGAGTÁGCATGAČAAAATCTT
AGAGCCCTCAGAGCAAAAATCCAGACATAGTTATCTATCAATATATGGATGACTTGTATGTAGGATCTGACTTAGAAA
TAGGGCAACATAGAGCAAAAATAGAAGAGTTAAGGAAACATTATTGAAATGGGATTACAACACCAGACAAGAAACAT
CAAAAAGAACCCCCATTCTTGGATGGGTATGAACCTCCATCTGACAAATGGACAGTACAACCTATACTGCTGCCAGA
AAAGGATAGTTGGACTGTCAATGATATACAGAAGTTAGTGGAAAATTAAACTGGCAAGTCAGATTACCCAGGGATT
AAGTAAGGCAACTCTGAACTCCTCAGGGGGCCAAGCACTAACAGACATAGTACCAACTGAGAAGCAGAATT
GAATTGGCAGAGAACAGGGAAATTAAAGAGAACAGTACATGGACTTATATTGATCCATCAAAGACTTGTATAGCTGA
AATACAGAAACAGGGCATGAAACATGGACATATCAAATTATCAAGAACCAATTAAAATCTGAAACAGGGAGTATG
CAAAAATGAGGACTACCCACACTAATGATGTTAACAGAGGGCAGTGCAAAAATAGCCATGGAAAGCATAGTA
ATATGGGGAAAGACTCTAAATTAGACTACCCATCCAAAAGAAACATGGGAGACATGGTGGACAGACTATTGCAAGC
CACCTGGATCCCTGAGTGGGAGTTGTTAACCCCTCCCCTAGTAAAATTATGGTACCAACTAGAAAAAGATCCATAG
CAGGAGTAGAAAATTCTATGTAGATGGGCAACTAACAGGAAAGCTAAAATAGGAAAAGCAGGGTATGTTACTGACAGA
GGAAGGGCAGAAAATTGTTACTCTAACTAACACAACAAATCAGAACAGACTGAGTTACAAGCAATTAGCTCTGCAGGA
TTCAGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCTAGGAATTCAGCTAGCTCTGCAGGA
CAGAGATTTAACCAATAATAGAACAGTTAATAAACAGGAAAGAATCTACCTGTATGGTACCCAGCACATAAAGGA
ATTGGGGAAATGAACAAGTAGATAAATTAGTAAGTAAGGAAATTAGGAAAGTGTG

FIGURE 42

>RevExon1_TV1_C_ZAopt (SEQ ID NO:72)

ATGGCCGGCCGCAGCGCGACAGCAGCGAGGCCCTGCTGCAGGTGGTGAAGATCATCAAGATCCTGTACCAGAGC

FIGURE 43

>RevExon1_TV1_C_ZAwt (SEQ ID NO:73)

ATGGCAGGAAGAACGGAGACAGCGACGAAGCGCTCCTCCAAGTGGTAAAGATCATCAAAATCCTCTATCAAAGCA

FIGURE 44

>RevExon2_TV1_C_ZAopt-2 (SEQ ID NO: 74)

CCCTACCCCAAGCCGAGGGCACCCGCCAGGCCGCAACCGCCGCCGCTGGCGGCCGCCAGCGCCAGATCCA
CACCATCGCGAGCGCATCTGGTGGCCTGCCTGGCCGCAGGCCGAGGCCGTGCCCCCTGCAGCTGCCCCCTGGAGC
GCCTGCACATCACTGCAGCGAGGGCAGCGGCACCAGCGGACCCAGCAGAGCCAGGGCACCACCGAGGGCGTGGCGAC
CCCTAA

FIGURE 45

>RevExon2_TV1_C_ZAwt (SEQ ID NO:75)

ACCCTTACCCCAAGCCCCAGGGGACTCGACAGGCTCGGAGGAATCGAAAGAAGAAGGTGGAGAGCAAGACAGAGACAGATC
CATACGATTGGTGAGCGGATTCTTGTGCGCTGGGACGATCTGCAGGCTGTGCCTCTTCAGCTACCACCGCTTGA
GAGACTTCATAATTAAATTGCAGTGAGGGCAGTGGAACTTCTGGGACACAGCAGTCTCAGGGGACTACAGAGGGGGTGGGAG
ATCCTTAA

FIGURE 46

RT_TV1_C_ZAopt (SEQ ID NO:76)

CCCATCAGCCCCATCGAGACCGTCCCCGTGAAGCTGAAGCCGGCATGGACGGCCCCA
AGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAAGATCAAGGCCCTGACCGCCATCTGCG
AGGAGATGGAGAAGGAGGGCAAGAGATCACCAAGATCGGCCCCGACAACCCCTACAACA
CCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTT
CCCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGCATCCCCCAC
CCCAGCCGGCCTGAAGAAGAAGAAGACCGTGTGGACGTGGCGACGCCCTAC
TTCAGCGTGCCCCCTGGACGAGAGCTCCGCAAGTACACCGCCTTACCATCCCCAGCA
TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCAAGGGCTGGAA
GGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTCCGCGCC
AAGAACCCCGACATCGTATCTACCAAGTACATGGACGACCTGTACGTGGCAGCGACC
TGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGT
GGGGCTTCACCACCCCCGACAAGAACCGACAGAGGAGCCCCCTTCTGTGGATGGG
CTACGAGCTGCACCCGACAAGTGGACCGTGCAGCCCATCCTGCTGCCCGAGAACCGAC
AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGCCAGCCAG
ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCCAAGGCC
TGACCGACATCGTCCCCCTGACCGAGGAGGCGAGCTGGAGCTGGCCAGAACCGCG
AGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCAGCAAGGACCTGATCGC
CGAGATCCAGAACGAGGGCCACGAGCAGTGGACCTACCAGATCTACCAGGAGCCCT
CAAGAACCTGAAGAACCGGCAAGTACGCCAAGATGCGCACCAACACCAACGACGT
GAAGCAGCTGACCGAGGCCGTGCAGAAGATGCCATGGAGAGACATCGTATCTGGGG
CAAGACCCCCAAGTTCCGCCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACC
GACTACTGGCAGGCCACCTGGATCCCCGAGTGGAGTTGTGAACACCCCCCCCCCTGG
TGAAGCTGTGGTACCAAGCTGGAGAACGGACCCATGCCGGCGTGGAGACCTTCTACGT
GGACGGCGCCACCAACCGCGAGGCCAAGATCGCAAGGCCGTACGTGACCGACCG
CGGCCGCCAGAACGATCGTGACCTGACCAACACCACCAACCGAGAACGACAGCTGCA
GGCATCCAGCTGGCCCTGCAGGACAGCGCAGCGAGGTGAACATCGTACCGACAG
CCAGTACGCCCTGGCATCATCCAGGCCAGCCGACAAGAGCGACAGCGAGATCTTC
AACCAAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGTGGCC
CCCACAAGGGCATGGCGGCAACGAGCAGGTGGACAAGCTGGTGGAGCAAGGGCATCC
GCAAGGTGCTG

FIGURE 47

>RT_TV1_C_ZAwT (SEQ ID NO:77)

CCAATTAGTCCTATTGAAACTGTACCACTAAAACCAGGAATGGATGGCCAAAGGTCAAACAATGCCATTGAC
AGAAGAAAAAATAAAAGCATTAAACAGCAATTGTGAGGAATGGAGAAGGAAGGAAAATTACAAAAATTGGCCTGATA
ATCCCATATAACACTCCAGTATTGCCATAAAAAGAAGGACAGTACTAAGTGAGAAAATTAGTAGATTCAAGGAAACTC
ATAAAAAGAACACTCAAGACTTTGGGAAGTTCAATTAGGAATACCCACACCCAGCAGGATTAAAAAGAAAAATCAGTGAC
AGTCTAGATGTGGGGATGCATAATTTCAGTTCTTAGATGAAAGCTTCAGGAATATACTGCATTACCCATACTA
GTATAAACATGAAACACCAGGGATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTC
CAGAGTAGCATGACAAAAATCTAGAGCCCTCAGAGCAAAAATCAGACATAGTTATCTATCAATATAATGGATGACTT
GTATGAGATCTGACTTAGAAATAGGGCAACATAGAGCAAAAATAGAAGAGTTAAGGGAACATTATTGAAATGGGAT
TTACAACACCAGACAAGAACATCAAAAGAACCCCCATTCTTGATGGGTATGAACTCCATCCTGACAATGGACA
GTACACCTATACTGCTGCCAGAAAAGGATAGTTGGACTGTCAATGATATAACAGAAGTTAGGGAAAATTAACCTGGC
AAGTCAGATTACCCAGGGATTAAAGTAAGGCAACTCTGTAACCTCTCAGGGGGCCAAAGCACTAACAGACATAGTAC
CACTACTGAAGAAGCAGAATTAGAATTGGCAGAGAACAGGGAAATTAAAGAGAACCCAGTACATGGAGTATATTGAT
CCATCAAAAGACTTGTAGCTGAAATACAGAAACAGGGCATGAACATGGACATATCAAATTATCAAGAACCATTTAA
AAATCTGAAAACAGGGAAAGTATGCAAAATGAGGACTACCCACACTAATGATGTAACAGTTAACAGAGGCAGTGCAA
AAATAGCCATGGAAAGCATAGTAATATGGGAAAGACTCCTAAATTAGACTACCCATCCAAAAGAACATGGAGACA
TGGTGGACAGACTATTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTGTTAATACCCCTCCCTAGTAAAATTATGGTA
CCAACATGAAAAGATCCCATAGCAGGAGTAGAAACTTCTATGTAGATGGAGCAACTAATAGGGAAAGCTAAAATAGGAA
AAGCAGGGTATGTTACTGACAGAGGAAGGAGAAAATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAA
GCAATTCACTAGCTCTGCAGGATTCAAGTAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCA
AGCACAACCAAGATAAGAGTGACTCAGAGATATTAAACCAATAATAGAACAGTTAACAAAGGAAAGAACATCTACCTGT
CATGGGTACAGCACATAAAGGAATTGGGGAAATGAACAAAGTAGATAAAATTAGTAAGTAAGGGAAATTAGGAAAGTGTG

FIGURE 48

>RTmut_TV1_C_ZAopt (SEQ ID NO: 78)

CCCATCAGCCCCATCGAGACCGTCCCCGTGAAGCTGAAGCCCCCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCGTGAC
CGAGGAGAAAGATCAAGGCCGTACCGCCATCTCGAGGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGACA
ACCCCTACAACACCCCCGTGTTGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGGAGCTG
AACAGCGCACCCAGGACTCTGGGAGGTGCAGCTGGCATCCCCCACCCCGCCGCTGAAGAAGAAGAAGAGCGTGAC
CGTGCCTGGACGTGGCGACGCCCTACTTCAGCGTCCCCCTGGAGGAGAGCTTCGCAGTACACCGCCCTCACCATCCCCA
GCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTCTGCCAGGGCTGGAGGGCAGCCCCGCCATCTC
CAGAGCAGCATGACCAAGATCCTGGAGCCCTCCGGCCAAGAACCCGACATCGTGAATCACCAGGCCCCCTGTACGT
GGCAGCGACCTGGAGATGCCAGCACCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGTGGGCTTCACCA
CCCCCGACAAGAACCGACCAAGGAGCCCCCTTCCTGCCATCGAGCTGCACCCGACAAGTGGACCGTGAGCCCAC
CTGCTGCCGAGAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACGGCCAGCCAGATCTA
CCCCGGCATCAAGGTGCGCAGCTGTGAAGCTGCTGCGGGGCCAAGGCCATCGGACATCGTGCCTGACCGAGG
AGGCCGAGCTGGAGCTGGCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCAGGGCTGTACTACGACCCAGCAAGGAC
CTGATCGCCGAGATCCAGAACGAGGGCACAGCAGTGGACCTACCAGATCTACCAAGGAGCCCTCAAGAACCTGAAGAC
CGGCAAGTACGCCAAGATGCCACCACACCAACGACGTGAAGCAGCTGACCGAGGGCGTGCAGAACGATGCCATGG
AGAGCATCGTATCTGGGCAAGACCCCAAGTCCGCTGCCATCCAGAAGGAGACCTGGAGACCTGGTGGACCGAC
TACTGGCAGGCCACCTGGATCCCCGAGTGGAGTTCTGAACACCCCCCTGGTGAAGCTGTGGTACCGAGCTGGAGAA
GGACCCCATGCCGGCGTGGAGACCTCTACGTGGACGGGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCTACG
TGACCGACCGCGGCCAGAAGATCGTGAACCTGACCAAACCCACCAACCGAGGCCAGTGCAGGCCATCCAGCTG
GCCCTGCAGGACAGCGGCCAGCGAGGTGAACATCGTACCGACAGCCAGTACGCCCTGGCATCATCCAGGCCAGCCGA
CAAGAGCGACAGCGAGATCTCAACCAGATCGACCGAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCCG
CCCACAAGGGCATCGCGGCCACGAGCAGGTGGACAAGCTGGTGGACAGGGCATCCGCAAGGTGCTG

FIGURE 49

>RTmut_TV1_C_ZAwT (SEQ ID NO: 79)

CCAATTAGTCCATTGAAACTGTACCAAGTAAAATTAAAACCAGGAATGGATGGCCCAAAGGTCAAACAATGCCATTGAC
AGAAGAAAAATAAAAGCATTAACAGCAATTGTGAGGAATGGAGAAGGAAGGAAAATTACAAAAATTGGCCTGATA
ATCCATATAACACTCCAGTATTGCCATAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAGGGAACTC
AATAAAAGAAGTCAAGACTTTGGGAAGTCAATTAGGAATACCACACCCAGCAGGATTAAAAAGAAAAATCAGTGAC
AGTGTAGATGTGGGGATGCATATTTTCAGTCCATTAGATGAAAGCTTCAGGAAATATACTGCAATTCAACCATACTA
GTATAAACAAATGAAACACCAGGGATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCGCAATATT
CAGAGTAGCATGACAAAATCTTAGAGCCCTTCAGAGCAAAAATCCAGACATAGTTATCTATCAAGCCCCCTGTATGT
AGGATCTGACTTAGAAATAGGGCAACATAGAGCAAAAATAGAAGAGTTAAGGGAACATTATTGAAATGGGATTTACAA
CACCAGACAAGAAACATCAAAAGAACCCCATTCTCCCATCGAACCTCCATGACAAATGGACAGTACAACCTATA
CTGCTGCCAGAAAGGATAGTTGGACTGTCAATGATATACAGAAGTTAGTGGAAAATTAAACTGGGCAAGTCAGATT
CCCAGGGATTAAAGTAAGGCAACTCTGAAACTCCTCAGGGGGCCAAGCACTAACAGACATAGTACCAACTGAAG
AAGCAGAAATTAGAATTGGCAGAGAACAGGGAAATTAAAGAGAACAGTACATGGAGTATTATGATCCATCAAAGAC
TTGATAGCTGAAATACAGAAACAGGGCATGAACAATGGACATATCAAGAACATTAAACTTGAAAAC
AGGAAAGTATGCAAAATGAGGACTACCCACACTAATGATGTTAACAGTTAACAGAGGCACTGCAAAAATAGCCATGG
AAAGCATAGTAATATGGGAAAGACTCTAAATTAGACTACCCATCCAAAAGAACATGGGAGACATGGTGGACAGAC
TATTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTGTTAATACCCCTCCCTAGTAAAATTATGGTACCAACTAGAAA
AGATCCCATAGCAGGAGTAGAAACTTCTATGTAGATGGCAACTAATAGGGAAAGCTAAAATAGGAAAAGCAGGGTATG
TTACTGACAGAGGAAGGCAGAAAATTGTTACTCTAACTAACACAACAAATCAGAACAGACTGAGTTACAAGCAATTCTAG
GCTCTGCAGGATTCAAGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCAATTAGGAATCTTCAAGCACAACCAG
TAAGAGTGAECTCAGAGATATTAAACAAATAATAGAACAGTTAATAAACAAGGAAAATCTACCTGTATGGTACCA
CACATAAGGAATTGGGGAAATGAACAAGTAGATAAATTAGTAAGTAAGGAAATTAGGAAAGTGTG
TTG

FIGURE 50

>TatC22Exon1_TV1_C_ZAopt (SEQ ID NO:80)

ATGGAGCCGTGGACCCAAGCTGAAGCCCTGGAACCACCCGGCAGCCAGCCAAGACC GCCGGCAACA ACTGCTTCTG
CAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCAGACCAAGGGCTGGCATCAGCTACGGCGCAAGAAGGCCGCC
AGCGCCGAGCGCCCCCCCCAGCGCGAGGACCACCAGAACCCCCCTGAGCAAGCAG

FIGURE 51

>TatExon1_TV1_C_ZAopt (SEQ ID NO:81)

ATGGAGCCCGTGGACCCAAGCTGAAGCCCTGGAACCACCCCGGCAGCCAGCCAAGACCGCCTGCAACAACACTGCTTCTG
CAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCCAGACCAAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGGCCGCC
AGCGCCCGAGCGCCCCCCCAGCGGCGAGGACCACCAGAACCCCCCTGAGCAAGCAG

FIGURE 52

>TatExon1_TV1_C_ZAwt (SEQ ID NO:82)

ATGGAGCCAGTAGATCCTAAACTAAAGCCCTGGAACCATCCAGGAAGCCAACCTAAAACAGCTTGTAAATAATTGCTTTG
CAAACACTGTAGCTATCATTGTCTAGTTGCTTCAGACAAAAGGTTAGGCATTCCTATGGCAGGAAGAAGCGGAGAC
AGCGACGAAGCGCTCCTCCAAGTGGTGAAGATCATCAAATCCTCTATCAAAGCAG

FIGURE 53

>TatExon2_TV1_C_ZAopt (SEQ ID NO:83)

CCCCCTGCCCCAGGCCCGCGCGACAGCACCGGCAGCGAGGAGAGCAAGAAGAAGGTGGAGAGCAAGACCGAGACCGACCC
CTACGACTGGTGA

FIGURE 54

>TatExon2_TV1_C_ZAwt (SEQ ID NO:84)

CCCTTACCCCAAGCCCGAGGGGACTCGACAGGCTGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGACAGAGACAGATCC
ATACGATTGGTGA

FIGURE 55

>Vif_TV1_C_ZAopt (SEQ ID NO:85)

ATGGAGAACCGCTGGCAGGTGCTGATCGTGTGGCAGGTGGACCGCATGAAGATCCGCCTGGAACAGCCTGGTGAAGCA
CCACATGTACATCAGCCGCCGCAGCGGCTGGGTGTACCGCCACCACTTCGAGAGCCGCCACCCCAAGGTGAGCAGCG
AGGTGCACATCCCCCTGGCGACGCCGCCTGGTGTACAGACCTACTGGGCCCTGCAGACCGCGAGCGCGACTGGCAC
CTGGGCCACGGCGTGAGCATCGAGTGGCGCTGCGCGAGTACAGCACCCAGGTGGACCCCGACCTGGCCGACCAGCTGAT
CCACATGCACTACTTCGACTGCTTCACCGAGAGCGCCATCCGCCAGGCATCCTGGGCCACATCGTGTCCCCCGCTGCG
ACTACCAGGCCGCCACAAGAAGGTGGGAGCCTGCAGTACCTGGCCCTGATCAAGCCAAGAAGCGCAAG
CCCCCCCCTGCCCAGCGTGCGCAAGCTGGTGGAGGACCGCTGGAACGACCCCCAGAAGACCCGGCGCCGGCAACCA
ACCATGAACGGCCACTAG

FIGURE 56

f_TV1_C_ZAwt (SEQ ID NO:86)

GGAAAACAGATGGCAGGTGCTGATTGTGTGGCAGGTGGACAGGATGAAGATTAGGCATGGAATAGTTAGTAAAGCA
TATGTATATATCAAGGAGAGCTAGTGGATGGGTCTACAGACATCATTTGAAAGCAGACATCCAAAAGTAAGTTCAG
TACATATCCATTAGGGATGCTAGATTAGTAATAAAACATATTGGGTTTGCAAGACAGGAGAAAGAGAATTGGCAT
GGTCATGGAGCTCCATAGAAATGGAGACTGAGAGAATACAGCACACAAGTAGACCCCTGACCTGGCAGACCAGCTAAT
CATGCATTATTGATTGTTTACAGAATCTGCCATAAGACAAGCCATATTAGGACACATAGTTTTCTTAGGTGTG
ATCAAGCAGGACATAAGAAGGTAGGATCTGCAATACTGGCACTGACAGCATTGATAAAACCAAAAAAGAGAAAG
CCTCTGCCTAGTGTAGAAAATTAGTAGAGGATAGATGGAACGACCCCCAGAAGACCAGGGGCCAGAGGGAACCA
AATGAATGGACACTAG

FIGURE 57

>vpr_TV1_C_ZAopt (SEQ ID NO: 87)

ATGGAGCGCCCCCCCAGGGACCAGGGCCCCCAGCGCGAGCCTACAACGAGTGGACCCCTGGAGATCCTGGAGGAGCTGAA
GCAGGAGGCCGTGCCTACCTCCCCCGCCCTGGCTGCACAGCCTGGCCAGTACATCTACGAGACCTACGGCGACACCT
GGACCGGGCGTGGAGGCCATCATCCCGGTGCTGCAGCAGCTGCTGTTCATCCACTTCCGCATCGGCTGCCAGCACAGCCGC
ATCGGCATCCTGCGCCAGCGCCGCCGCAACGGCGCCAGCCGCAGC

FIGURE 58

>Vpr_TV1_C_ZAwT (SEQ ID NO: 88)

ATGGAACGACCCCCAGAAGACCAGGGGCCGAGAGGGAACCATACAATGAATGGACACTAGAGATTCTAGAAGAACTCAA
GCAGGAAGCTGTCAGACACTTCCCTAGACCATGGCTCCATAGCTTAGGACAATATATCTATGAAACCTATGGGATACTT
GGACGGGAGTTGAAGCTATAATAAGAGTACTGCAACAACTACTGTTCATTCAATTCAAGAATTGGATGCCAACATAGCAGA
ATAGGCATTTGCGACAGAGAAGAGCAAGAAATGGAGCCAGTAGATCC

FIGURE 59

>Vpu_TV1_C_ZAopt (SEQ ID NO:89)

ATGGTGAGCCTGAGCCTGTTCAAGGGCGTGGACTACCGCCTGGCGTGGGCCCTGATCGTGGCCCTGATCATGCCAT
CATCGTGTGGACCACATCGCCTACATCGAGTACCGCAAGCTGGTGCAGAAGAAAGATCGACTGGCTGATCAAGGCCATCC
GCGAGCGCGCCGAGGACAGCGGAAACGAGAGCGACGGGACACCGAGGAGCTGAGCACCATTGGTGGACATGGGCCACCTG
CGCCTGCTGGACGCCAACGACCTGTAA

FIGURE 60

>Vpu_TV1_C_ZAw (SEQ ID NO:90)

ATGGTAAGTTAACGAGTAGATTATAGATTAGGAGTAGGAGCATTGATAGTAGCACTAATCATAGCAAT
AATAGTGTGGACCATAGCATATATAGAATATAGGAAATTGGAAGACAAAAGAAAATAGACTGGTTAATTAAAAGAATTA
GGGAAAGAGCAGAACAGTGCGAATGAGAGTGATGGGACACAGAAGAATTGTCACAATGGTGGATATGGGCATCTT
AGGCTTCTGGATGCTAATGATTGTAA

FIGURE 61

dna revexon1_2TV1_C_ZAop (SEQ ID NO:91)

ATGGCCGGCCGCAGCGGCGACAGCGACGAGGCCCTGCTGCAGGTGGTGAAGATCATC
AAGATCCTGTACCAGAGCCCCTACCCCAAGCCCGAGGGCACCCGCCAGGCCGCCGCA
ACCGCCGCCGCCGCTGGCGCGCCGCCAGCGCCAGATCCACACCATGGCGAGCGCAT
CCTGGTGGCCTGCCTGGGCCGCAGCGCCGAGCCCGTGCCTGCAGCTGCCCTG
GAGCGCCTGCACATCACTGCAGCGAGGGCAGCGGACCCAGCGGACCCAGCAGAGC
CAGGGCACCACCGAGGGCGTGGCGACCCCTAA

FIGURE 62

dna Revexon1_2_TV1_C_ZAwt (SEQ ID NO:92)

ATGGCAGGAAGAACGGAGACAGCGACGAAGCGCTCTCCAAGTGGTGAAGATCATC
AAAATCCTCTATCAAAGCAACCCTAACCCCAAGCCCAGGGGACTCGACAGGCTCGGA
GGAATCGAAGAAGAAGGTGGAGAGCAAGACAGAGACAGATCCATACGATTGGTGAGC
GGATTCTTGTGCGCTGGGACGATCTGCAGCCTGTGCCTCTTCAGCTACCACCG
CTTGAGAGACTTCATATTAAATTGCAGTGAGGGCAGTGGAACTTCTGGGACACAGCAGT
CTCAGGGGACTACAGAGGGGTGGGAGATCCTAA

FIGURE 63

dna TatC22Exon1_2_TV1_C_ZAopt (SEQ ID NO:93)

ATGGAGCCCGTGGACCCAAGCTGAAGCCCTGGAACCACCCGGCAGCCAGCCCAAG
ACCGCCGGCAACAACACTGCTTCTGCAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCC
AGACCAAGGGCCTGGCATCAGCTACGGCCGCAAGAACGGGAGGACCAAGAACCCCTGAGCAAGCAGCCCCCTGCCAGGC
CCCCCCCCAGCGCGAGGACCAAGAACCCCTGAGCAAGAACGGGAGGAGCAAGAACGG
CCGCGCGACAGCACCGGCAGCGAGGAGAGCAAGAACGGTGGAGAGCAAGACCG
AGACCGACCCCTACGACTGGTGA

FIGURE 64

dna TatExon1_2_TV1_C_ZAopt (SEQ ID NO:94)

ATGGAGCCGTGGACCCAAGCTGAAGCCCTGGAACCACCCCGGCAGCCAGCCCAAG
ACCGCCTGCAACAACACTGCTTCTGCAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCCA
GACCAAGGGCTGGCATCAGCTACGGCCGCAAGAACCGGCTGAGCAAGCAGCCCCCTGCCAGGGCCC
CCCCCCAGCGCGAGGACCACAGAACCGGCTGAGCAAGAACGGTGGAGAGCAAGACCGAG
GCGCGACAGCACCGGCAGCGAGGAGAGCAAGAACGGTGGAGAGCAAGACCGAG
ACCGACCCCTACGACTGGTGA

FIGURE 65

dna TatExon1_2_TV1_C_ZAwT (SEQ ID NO:95)

ATGGAGCCAGTAGTCCTAAACTAAAGCCCTGGAACCATCCAGGAAGCCAACCTAAA
ACAGCTTGTATAATTGCTTTGCAAACACTGTAGCTATCATTGTCTAGTTGCTTTCA
GACAAAAGGTTAGGCATTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGC
TCCTCCAAGTGGTGAAGATCATAAAATCCTATCAAAGCAGCCCTACCCCAAGCC
CGAGGGGACTCGACAGGCTCGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGACAGA
GACAGATCCATACGATTGGTGA

FIGURE 66

NefD125G-Myr_TV1_C_ZAopt (SEQ ID NO:96)

ATGGCCGGCAAGTGGAGCAAGCGCAGCATCGTGGGCTGGCCCGCCGTGC
GAGCGCATGCGCCGCACCGAGCCGCCGAGGGCGTGGCGCCGCCAGC
CAGGACCTGGACC GCCACGGCGCCCTGACCAGCAGCAACACCCCCGCCACCA
ACGAGGCCTGCGCCTGGCTGCAGGCCAGGAGGAGGACGGCACGTGGGCT
TCCCCGTGCGCCCCCAGGTGCCCCTGCGCCCCATGACCTACAAGAGCGCCGT
GGACCTGAGCTTCTCCTGAAGGAGAAGGGCGGCCTGGAGGGCTGATCTAC
AGCCGCAAGCGCCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGCT
TCTTCCCCGGCTGGCAGAACTACACCAGCGCCCCGGCGTGCCTTCCCCCTG
ACCTTCGGCTGGTGCTCAAGCTGGTGCCTGGACCCCCCGAGGTGAAGG
AGGCCAACGAGGGCGAGGACAAC TGCCCTGCTGCACCCCATGAGCCAGCACG
GCGCCGAGGACGAGGACCGCGAGGTGCTGAAGTGGAAAGTTCGACAGCCTGC
TGGCCCACCGCCACATGGCCCGCGAGCTGCACCCCGAGTACTACAAGGACTG
CTGA

FIGURE 67

ATGCGGCCGGCATCCTGAAGAACTACCGCCACTGGTGGATCTGGGCATCCT
GGCTCTGGATGCTGATGTCAACGTGAAGGGCCTGTGGGTGACCGTGTACTA
CGCGTCCCCGTGGGCCGAGGCCAAGACCACCCCTGTTCTGCGCCAGCGACGCCA
AGGCCTACGAGAAGGAGGTGCACAACGTGTGGGCCACCCACGCCCTGCGTCCCCACC
GACCCCAACCCCCAGGAGGTGATCCTGGCAACGTGACCGAGAACTTCAACATGTG
GAAGAACGACATGGTGGACCAGATGCAGGAGGACATCATCAGCCTGTGGGACCAGA
GCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTGACCCCTGAACTGCACCAACG
CCACCGTGAACTACAACAACCCAGCAAGGACATGAAGAACTGCAGCTTACGTG
ACCACCGAGCTGCGGACAAGAAGAAGAAGGAGAACGCCCTGTTCTACCGCTGG
CATCGTCCCCCTGAACAACCGCAAGAACGGCAACATCAACAACACTACCGCTGATCA
ACTGCAACACCAAGGCCATACCCAGGCCCTGCCCCAAGGTGAGCTTCGACCCCATCC
CCATCCACTACTGCGCCCCCGCCGGCTACGCCCTGAAGTGCAACAACAAGAAG
TTCAACGGCATCGGCCCTGCGACAACGTGAGCACCGTGCAGTGACCCACGGCAT
CAAGCCCGTGGTGAGCACCCAGCTGCTGTAACGGCAGCCTGGCGAGGAGGAGA
TCATCATCCGCAGCGAGAACCTGACCAACAACGTGAAGGACCATCATCGTGCACCTG
AACGAGAGCATCGAGATCAAGTGACCCGCCGGCAACAACACCCGCAAGAGCGT
GCGCATCGGCCCCGGCAGGCCTTCTACGCCACCGCGACATCATCGCGACATCC
GCCAGGCCACTGCAACATCAGCAAGAACGAGTGGAACACCCACCTGCAGCGCTG
AGCCAGAAGCTGCAGGAGCTGTTCCCAACAGCACCGCATCAAGTTCGCCCCCA
CAGCGCGGGGACCTGGAGATCACCACCCACAGCTCAACTGCGGGAGTTCT
TCTACTGCAACACCAACCGACCTGTTCAACAGCACCTACAGCAACGGCACCTGCACCA
ACGGCACCTGCATGAGCAACAACACCGAGCGCATCACCTGCAGTGCACATCAAG
CAGATCATCAACATGTGGCAGGAGGTGGCCGCGCATGTCAGCCCCCCCCATCGC
CGGCAACATCACCTGCCGAGCAACATCACCGGCGTGTGTCACCGCGACGGCG
GCGACAACAACACCGAGACCGAGACCTCCGCCGGCGGCGACATGCGCGAC
AACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGAGATCAAGCCCTGGCGT
GGCCCCCACCGCCGCAAGCGCCGCGTGGAGCGCGAGAACGCGCCGGCGAC
TCGGCGCCGTGTTCTGGCTTCTGGCGCCGCGCGCAGCACCATTGGCGCCGCA
GCATCACCTGACCGTGCAGGCCATCGAGGCCAGCAGCACATGTCAGCTGACCGTGTG
AGCAACCTGCTGCGGCCATCGAGGCCAGCAGCACATGTCAGCTGACCGTGTG
GGGCATCAAGCAGCTGCAGGCCCGCGTGTGGCCATCGAGCGCTACCTGCAGGACC
AGCAGCTGCTGGCCTGTGGGCTGCAGCGCAAGCTGATCTGCACCAACGTG
CTGTGGAACAGCAGCTGGAGCAACAAGACCCAGAGCGACATCTGGGACAACATGAC
CTGGATGCAGTGGACCGCGAGATCAGCAACTACACCAACCCATCTACCGCTGC
TGGAGGACAGCCAGAGCCAGCAGGAGCGCAACGAGAACGGACCTGCTGGCCTGGA
CCGCTGGAACAACCTGTGGAACTGGTTCAGCATACCAACTGGCTGTGGTACATCAA
GATCTTCATCATGATCGTGGCGGCCCTGATCGGCCATCGACATCTTCGCCGTGCT
GAGCCTGGTGAACCGCGTGCAGGCCAGGGCTACAGCCCCCTGAGCCTGCAGACCCCTGA
TCCCCAACCCCGCGGCCCGACCGCTGGCGCATCGAGGAGGAGGGCGGCGAG
CAGGACAGCAGCCGAGCATCCGCTGGTGAAGCGGCTTCTGACCCCTGGCCTGGGA
CGACCTGCGCAGCCTGTGCCTGTTCTGCTACCAACCGCCTGCGGACTTCATCCTGAT
CGTGGTGCAGCCGCGTGGAGCTGCTGGGCCACAGCAGCCTGCGCGCCTGCAGCGCG
GCTGGGGCACCCCTGAAGTACCTGGCAGCCTGGTGCAGTACTGGGGCCTGGAGCTG
AAGAAGAGCGCCATCAACCTGCTGGACACCATGCCATGCCGTGGCGAGGGCAC
CGACCGCATCCTGGAGTTCATCCAGAACCTGTGCCGCGCATCCGCAACGTGCCCG
CCGCATCCGCCAGGGCTTCGAGGCCGCCCTGCAGTAA

FIGURE 68

ATGAGAGCGAGGGGGATACTGAAGAATTATCGACACTGGTGGATATGGGCATCTT
AGGCTTTGGATGCTAATGATGTGAAATGTGAAGGGCTGTGGTCACAGTCTACTA
CGGGGTACCTGTGGGGAGAGAACGAAAAACTACTCTATTTGTGCATCAGATGCTA
AAGCATATGAGAAAGAAGTCATAATGTCTGGCTACACATGCCGTGTACCCACA
GACCCCACCCACAAGAAGTGATTTGGCAATGTAACAGAAAATTAAACATGTG
GAAAAATGACATGGTGGATCAGATGCAGGAAGATATAATCAGTTATGGATCAA
GCCTTAAGCCATGTGTAATTGACCCACTCTGTGTCACTTAAACTGTACAAATG
CAACTGTTAACTACAATAATACCTCTAAAGACATGAAAAATTGCTCTTCTATGTAA
CCACAGAATTAAAGAGATAAGAAAAAGAAAATGCACTTTTATAGAATTGAT
ATAGTACCACTTAATAATAGGAAGAATGGGAATTAAACAACATATAGATTAAATAA
TTGTAATACCTCAGCCATAACACACAGCCTGCCAAAAGTCCTGTTGACCCAATTCC
TATACATTATTGTGCTCCAGCTGGTTATGCGCCTCTAAATGTAATAATAAGAAAATT
CAATGGAATAGGACCATGCGATAATGTCAGCACAGTACAATGTACACATGGAATT
AGCCAGTGGTATCAACTCAATTACTGTTAAATGGTAGCCTAGCAGAAGAAGAGATA
ATAATTAGATCTGAAAATCTGACAAACAATGTCAAAACAATAATAGTACATCTTAAT
GAATCTATAGAGATTAAATGTACAAGACCTGGCAATAATACAAGAAAGAGTGTGAG
AATAGGACCAAGGACAAGCATTCTATGCAACAGGAGACATAATAGGAGATATAAGAC
AAGCACATTGTAACATTAGTAAAATGAATGGAATACAACATTACAAAGGTAAGT
CAAAATTACAAGAACTCTCCCTAATAGTACAGGGATAAAATTGCAACCACACTCA
GGAGGGGACCTAGAAATTACTACACATAGCTTAATTGTGGAGGAGAATTCTAT
TGCAATACAACAGACCTGTTAATAGTACATACAGTAATGGTACATGCACTAATGGT
ACATGCATGTCTAATAATACAGAGCGCATCACACTCCAATGCAAGATAAAACAAAT
TATAAACATGTGGCAGGAGGTAGGACGAGCAATGTATGCCCTCCATTGCAGGAA
ACATAACATGTAGATCAAATTACAGGACTACTATTAAACACGTGATGGAGGAGAT
AATAACTGAAACAGAGACATTCAAGACCTGGAGGAGACATGAGGGACAATTG
GAGAAGTGAATTATATAAATACAAGGTGGTAGAAATTAAACCATTAGGAGTAGCAC
CCACTGCTGCAAAAGGAGAGTGGTGGAGAGAGAAAAAGAGCAGTAGGAATAGG
AGCTGTGTTCTGGGTTCTGGGAGCAGCAGGAAGCACTATGGCGCAGCATCAAT
AACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCACAGCAAAGTA
ATTGCTGAGGGCTATAGAGGCACAGCATATGTTCAACTCACGGCTGGGC
ATTAAGCAGCTCCAGGCAAGAGTCCTGGCTATAGAGAGATACTACAGGATCAACA
GCTCCTAGGACTGTGGGCTGCTCTGGAAAACCTCATGCAACCAACTAATGTGCTTG
GAACCTAGITGGAGTAATAAAACTCAAAGTGTATTTGGGATAACATGACCTGGAT
GCAGTGGGATAGGGAAATTAGTAATTACACAACACAATACAGGTGCTGAAG
ACTCGCAAAGCCAGCAGGAAGAAAGAAAAAGATTACTAGCATTGGACAGGTGG
AACAACTGTGGAATTGGTTAGCATAACAAATTGGCTGTGGTATATAAAATATT
ATAATGATAGTAGGAGGCTGATAGGTTAAGAATAATTGTGCTGTCTCT
GTAAATAGAGTTAGGCAGGGATACTCACCCCTGTCTGGCAGACCCCTATCCCAAAC
CCGAGGGGACCGACAGGCTGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGACA
GCAGCAGATCCATTGATTAGTGGAGCGGATTCTGACACTTGCCGTGGACACTAC
GAAGCCTGTGCCTCTCTGCTACCAACCGATTGAGAGACTTCATATTAAATTGTAGTGA
GAGCAGTGGAACTTCTGGGACACAGTAGTCTCAGGGGACTGCAGAGGGGTGGGA
ACCCCTAAGTATTGGGAGTCTTGCAATATTGGGTCTAGAGTTAAAAAAGAGT
GCTATTAACTGCTGATACTATAGCAATAGCAGTAGCTGAAGGAACAGATAGGATT
CTAGAATTACACAAACCTTGTAGAGGTATCCGCAACGTACCTAGAAGAATAAG
ACAGGGCTCGAAGCAGCTTGCAATAA

FIGURE 69

Gag_TV2_C_ZAopt (SEQ ID NO:99)

ATGGGCGCCGCCAGCATCCTGCGCGGCGGAAGCTGGACAAGTGGAG
AAGATCCGCCTGCGCCCCGGCGCGCAAGCACTACATGCTGAAGCACCTGG
TGTGGGCCAGCCGCGAGCTGGAGCGCTCGCCGTGAACCCCGGCTGCTGG
GACCAGCGACGGCTGCCGCCAGATCATCAAGCAGCTGCAGCCCCCCCCTGCAG
ACCGGCACCGAGGAGATCCGCAGCCTGTTCAACACCGTGGCCACCCCTGTACT
GCGTCACAAGGGCATCGACGTGCGCGACACCAAGGAGGCCCTGGACAAGA
TCGAGGAGGAGCAGAACAAAGTGCCAGCAGAAGACCCAGCAGGCCAGGCG
CCGACAAGAAGGTGAGCCAGAACTACCCCATCGTGAGAACCTGCAGGGCC
AGATGGTGCACCAGGCCATCAGCCCCCGCACCTGAACGCCCTGGTGAAGGT
GATCGAGGAGAACGGCCTTCAGCCCCGAGGTGATCCCCATGTTCACCGCCCTG
AGCGAGGGCGCCACCCCCCAGGACCTGAACACCATGCTGAACACCGTGGGC
GCCACCAAGGCCATGAGATGCTGAAGGACACCATAACGAGGAGGCC
GCCGAGTGGGACCGCCTGCACCCGTGCACGCCGGCCCCGTGGCCCCCGGCC
AGATGCGCGAGCCCCCGGGCAGCGACATCGCCGGCACCAAGCACCCTGCA
GGAGCAGATCGCCTGGATGACCGAGCAACCCCCCATCCCCGTGGCGACATC
TACAAGCGCTGGATCATCCTGGCCTGAACAAGATCGCGCATGTACAGCC
CCGTGAGCATCCTGGACATCAAGCAGGGCCCCAAGGAGGCCCTCCGCGACTA
CGTGGACCGCCTTCAGAACCCCTGCGCGCCAGCAGAGCACCCAGGAGGTG
AAGAACTGGATGACCGACACCCCTGCTGGTGAGAACGCCAACCCGACTGCA
AGACCATCCTGCGGCCCTGGCCCCGGGCCAGCCTGGAGGAGATGAC
CGCCTGCCAGGGCGTGGCGGCCAGCCACAAGGCCCGTGTGGCCAG
GCCATGAGCCAGGCCACAACACCAGCGTGTGATCCAGAAGAGCAACTTC
AAGGGCCCCCGCCGCCGTGAAGTGCTCAACTGCGGCCGCAGGGCCACA
TCGCCCCGCAACTGCCGCCCTGGCAAGCGCGGCTGTGGAAAGTGC GGCAA
GGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTGG
CAAGATCTGGCCCAGCCACAAGGGCCCCCCGGCAACTTCTGAGGCCG
CCCGAGGCCACGCCCTGGAGGCCACCGGCCGGCGAGAGCT
TCAAGTTCAAGGAGACCCCCAAGCAGGAGGCCAAGGAGCCGAGGCCCTGA
CCAGCCTGAAGAGCCTGTTCGGCAGCGACCCCTGAGGCCAGTAA

FIGURE 70

Gag_TV2_C_ZAwt (SEQ ID NO:100)

ATGGGTGCGAGAGCGTCAATATTAAAGAGGGGGAAAATTAGACAAATGGGAA
AAAATTAGGTTACGGCCAGGGGGAGAAAACACTATATGCTAAAACACCTA
GTATGGGCAAGCAGAGAGCTGGAAAGATTGCAGTTAACCTGGCCTTTAG
AGACATCAGACGGATGTAGACAAATAATAAAACAGCTACAACCAGCTCTCA
GACAGGAACAGAGGAATTAGATCATTATTAAACACAGTAGCAACTCTCTAT
TGTGTACATAAAGGGATAGATGTACGAGACCCAAGGAAGCCTAGACAG
ATAGAGGAGGAACAAAACAAATGTCAGCAAAAAACACAGCAGGCCAGCG
GCTGACAAAAAGGTCACTCAAATTATCCTATAGTCAGAACCTCCAAGGGC
AAATGGTACACCAGGCCATATCACCTAGAACCTTGAATGCATGGTAAAAGT
AATAGAGGAGGAAGGCTTTAGCCAGAGGTAATAACCCATGTTACAGCATT
TCAGAAGGAGCCACCCCACAAGATTAAACACCATGTTAATACAGTGGGG
GACATCAAGCAGCCATGCAAATGTTAAAAGATACCATCAATGAGGAGGCTGC
AGAATGGGATAGGTTACATCCAGTACATGCAGGGCTGTTGCACCAGGCCAG
ATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACACTAGTACCCCTCAA
GAACAAATAGCATGGATGACAAGTAACCCACCTATCCCAGTAGGGGACATCT
ATAAAAGGTGGATAATTCTGGGTTAAATAAAATAGTAAGAATGTACAGCCC
TGTCAAGCATTAGACATAAAACAAGGACCAAAAGGAACCCCTTAGAGACTAT
GTAGACCGGTTCTCAAACITTAAGAGCTGAACAATCTACACAAGAGGTAA
AAAATTGGATGACAGACACCTTGTAGTCCAAAATGCGAACCCAGATTGTA
GACCATTAAAGAGCATTAGGACCAAGGGCTTCATTAGAAGAAATGATGACA
GCATGTCAGGGAGTGGGAGGACCTAGCCACAAAGCAAGAGTTGGCTGAG
GCAATGAGCCAAGCAAACAATACAAGTGTAAATGATAACAGAAAAGCAATTAA
AAGGCCCTAGAAGAGCTGTTAAATGTTCAACTGTGGCAGGGAAAGGGCACAT
AGCCAGGAATTGCAGGGCCCTAGGAAAAGGGGCTGTTGAAATGTGGAAA
GGAAGGACACCAAATGAAAGACTGTACTGAGAGGCAGGCTAATTAGGG
AAAATTGGCCTTCCCACAAGGGGAGGCCAGGGAAATTCCCTCAGAGCAGAC
CAGAGCCAACAGCCCCACCACTAGAACCAACAGCCCCACCAAGCAGAGGCT
TCAAGITCAAGGAGACTCCGAAGCAGGAGCCGAAAGACAGGGAACCTTAA
CTTCCCTCAAATCACTTTGGCAGCGACCCCTGTCTCAATAA

FIGURE 71

Nef_TV2_C_ZAopt (SEQ ID NO:101)

ATGGCGGGCAAGTGGAGCAAGAGCAGCATCGGCTGGCCCGAGGTGCGC
GAGCGCATCCGCCGCACCGCAGCGCCGCCAGGGCGTGGCAGCGCCAGC
CAGGACCTGGAGAACGACGGCGCCCTGACCACCAGCAACACCGCCCACAAAC
AACGCCGCCCTGCGCCTGGCTGGAGGCCAGGAGGAGGAGGGCGAGGTGGC
TTCCCCGTGCGCCCCCAGGTGCCCCCTGCGCCCCATGACCTACAAGGCCGCAT
CGACCTGAGCTTCTTCCTGAAGGAGAACGGCGGCCTGGAGGGCCTGATCTAC
AGCAAGAACGCCAGGAGATCTGGACCTGTGGGTGTACAACACCCAGGGC
TTCTTCCCCGACTGGCAGAACTACACCCCCGGCCCCGGCGTGCCTCCCCCT
GACCTTCGGCTGGTACTCAAGCTGGAGCCCGTGGACCCCCCGCGAGGTGGAG
GAGGCCAACGAGGGCGAGAACAACTGCCTGCTGCACCCATGAGCCAGCAC
GGCATGGAGGACGAGGACCGCGAGGTGCTGCGCTGGAAGTTGACAGCACC
CTGGCCCGCCACATGGCCCGAGCTGCACCCGAGTAACAGGACT
GCTGA

FIGURE 72

Nef_TV2_C_ZAwt (SEQ ID NO:102)

ATGGGGGGCAAGTGGTCAAAAGCAGTATAATTGGATGGCCTGAAGTAAGA
GAAAGAACATCAGACGAACTAGGTAGCAGCAGAGGGAGTAGGATCAGCGTCT
CAAGACTTAGAGAAACATGGGCACTTACAACCAGCAACACAGCCCACAAC
AATGCTGCTTGCCTGGCTGGAAGCGCAAGAGGGAGGAAGGAGAAGTAGGC
TTCCAGTCAGACCTCAGGTACCTTAAGACCAATGACTTATAAAGCAGCAAT
AGATCTCAGCTTCTTTAAAAGAAAAGGGGGACTGGAAGGGTTAATTAC
TCCAAGAAAAGGCAAGAGAGATCCTGATTGTGGTTATAACACACAAGGCT
TCTCCCTGATTGGCAAAACTACACACCCGGGACCAGGGTCAGATTCCACT
GACCTTGGATGGTACTCAAGCTAGAGCCAGTCGATCCAAGGGAAAGTAGAA
GAGGCCAATGAAGGAGAAAACAACACTGTTACTACACCCTATGAGCCAGCATG
GAATGGAGGATGAAGACAGAGAAGTATTAAGATGGAAGTTGACAGTACGC
TAGCACCGCAGACACATGGCCCGAGCTACATCCGGAGTATTACAAAGACTG
CTGA

FIGURE 73

Pol_TV2_C_ZAopt (SEQ ID NO:103)

TTCTTCGCGAGAACCTGGCTTCCCCAGGGCGAGGCCCGAGTTCCCCAGCGAGCAGACC
CGCGCCAACAGCCCCACCAACCGCACCAACAGCCCCACCAGCCGAGCTGCAGGTGCAGGG
CGACAGCAGGCCGCCAGCGCCAGGGCACCTCAACTTCCCCAGATCACCCGTGGC
AGCGCCCCCTGGTGAGCATCAAGGTGGCCGGCAGACCAAGGAGGCCCTGCTGGACACCCGC
GCCGACGACACCGTGCTGGAGGAGATCAACCTGCCGGCAAGTGGAAAGCCAAGATGATCGG
CGGCATCGGCGGCTCATCAAGGTGCAGTACGACCAGATCCTGATCGAGATCTGCAGCA
AGCGGCCATCGGCACCGTGTGGGGCCCCACCCCGTGAACATCATCGGGCGAACCTGC
TGACCCAGCTGGGCTGCACCCCTGAACCTCCCCATCAGCCCCATCGAGACCCTGCCCCGTGAAGC
TGAAGCCGGATGGACGGCCCCAAGGTGAAGCAGTGGGGCTGACCGAGGGAGAAGATCAAG
GCCCTGACCGAGATCTGCAGGGAGATGGAGAAGGAGGGCAAGATACCAAGATCGGCCCCG
AGAACCCCTACAACACCCCCGTGTCGCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG
CTGGTGGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCAT
CCCCCACCCGCCGGCCTGAAGAAGAAGAAGAGCGTGCACCGTGTGGACGTGGCGACGCC
ACTTCAGCGTGCCTGGACGAGAGCTCCGCAAGTACACCGCCTCACCATCCCCAGCATCA
ACAACGGAGACCCCGGCATCCGCTACCAAGTACAACCGTGTGCCCAAGGGCTGGAAGGGAGC
CCCGCCATCTTCCAGAGCAGCATGACCCGCATCTGGAGCCCTCCGCACCCAGAACCCCGAG
GTGGTGTACTACCACTACATGGACGACCTGTACGTGGCAGCGACCTGGAGATCGGCCAGCA
CCGCGCCAAGATCGAGGGAGCTGCACGGGCCACCTGCTGAAGTGGGCTTACCAACCCCCGACA
AGAACGACCAAGAAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCACCCAGACAAGTGG
ACCGTGCAGCCCATCCAGCTGCCAGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCT
GGTGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCAGCTGTGCA
AGCTGCTGCACGGCGCAAGGCCCTGACCGACATCGTGCCTGACCGAGGGCGAGCTG
GAGCTGGCCGAGAACCGCGAGATCTGAAGGAGCCGTGACGGCGTGTACTACGACCCAG
CAAGGACCTGATCGCCAGATCCAGAACGAGGGCAACGACCAAGTGGACCTACCAAGATCTACC
AGGAGCCCTCAAGAACCTGCGCACCGCAAGTACGCCAGATGCGCACCGCCACACCAAC
GACGTGAAGCAGCTGGCCGAGGCCGTGACGGAGATCACCCAGGAAGCATCGTGTACTGGGG
CAAGACCCCAAGTCCGCTGCCCATCCCCAAGGAGACCTGGGAGACCTGGTGGAGCGACT
ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCTGTGAACACCCCCCCCCTGGTGAAGCTGT
GGTACCACTGGAGAAGGAGCCCATCGTGGCGCCGAGACCTTCTACGTGGACGGCGCC
AACCGCGAGACCAAGATCGCAAGGCCGCTACGTGACCGACAAAGGGCGCCAGAAGGTGG
TGAGCTTCAACCGAGACCAACCAGAACGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAG
GACAGCGGGCCCGAGGTGAACATCGTGCACGCCAGTACGCCCTGGCATCATCCAGGC
CCAGCCCAGAACAGAGCGAGAGCGAGCTGGTGAAGCCAGATCATCGAGCAGCTGATCAAGAAG
GAGAAGGTGTACCTGAGCTGGGTGCCGCCACAAGGGCATGGCGCAACGAGCAGGTGGA
CAAGCTGGTAGCAGCGGCATCCGAAGGTGTGTTCTGGACGGCATCGACAAGGCCAGG
AGGAGCACGAGAACGACCAACTGGCGGCCATGGCAGCGAGTTCAACCTGCC
ATCGTGGCCAAGGAGATCGTGGCCAGCTGCAGCAAGTGCAGCTGAAGGGCGAGGCCATGCA
CGGCCAGGTGACTGCAGCCCCGGCATCTGGCAGCTGACTGCACCCACCTGGAGGGCAAGA
TCATCCTGGTGGCCGTGCACTGGCCAGGCCATCGTGCAGGCCAGGAGGTGATCCCCGCCAG
ACCGGCCAGGAGACCGCTACTTCATCCTGAAGCTGGCGGCCGTGGCCCGTGAAGGTGATC
CACACCGACAACGGCAGCAACTTCAACCGACCCGCGTGAAGGCCGCTGCTGGTGGCCGA
CATCCAGCGCGAGTTGGCATCCCCACAACCCAGAGCCAGGGCGTGGTGGAGAGCATGA
ACAAGGAGCTGAAGAACGATCATGGCCAGGTGCGCACCGAGGCCAGCCTGAAGACCGCC
GTGCAGATGGCCGTGTTCATCCACAACCTCAAGCGCAAGGGCGCATCGGCCGCTACAGCGC
CGGCCAGCGCATCATCGACATCATCGCAGCGACATCCAGACCAAGGAGCTGCAGAACGAGA
TCATCAAGATCCAGAACCTGGCGTGTACTACCGCGACAGCCGCAACCCATCTGGAAAGGGCC
CCGCCAAGCTGCTGTGGAAGGGCGAGGGCGCCGTGGTATCCAGGACAACAGCGACATCAAG
GTGGTGGCCCGCCAGGCAAGGCAAGATCATCAAGGACTACGGCAAGCAGATGGCGGCCGA
CTCGTGGCCGGCCAGGACGAGGAC

FIGURE 74

Pol_TV2_C_ZAwt (SEQ ID NO:104)

TTTTTAGGGAAAATTGGCCTTCCCACAAGGGGAGGCCAGGAATTCCCTCAGAGCAGACC
AGAGCCAACAGCCCCACCACTAGAACCAACAGCCCCACCAGCAGAGAGCTCAAGTCAGG
AGACTCCGAACAGCAGGAGCGAAAGACAGGGAACCTTAACCTCCCTCAAATCACTCTTGCA
GCGACCCCTTGCTCTCAATAAAAGTAGCGGGCCAAACAAAGGAGGCTTTAGATACAGGAG
CAGATGATACAGTACTAGAAGAAATAACTGCCAGGGAAATGGAACCAAAAAATGATAGG
AGGAATTGGAGGTITATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTGTGGAAA
AAGGGCTATAAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTT
GACTCAGCTTGATGCACACTAAATTCCATTAGCCCCATTGAAACTGTACAGTAAAATT
AAAGCCAGGAATGGATGGCCAAAGGTTAACAAATGCCATTGACAGAAGAAGAAGAAGAAGA
GCATAACAGAAATTGTGAGGAAATGGAGAAGGAAAGGAAATTACAAAAATTGGCCTGA
AAATCCATATAACACTCCAGTATTGCCATAAGAAGAAGGACAGTACAAAGTGGAGAAAAT
TAGTAGATTTCAGGAACTCAATAAAAGAACTCAAGACTTTGGAAAGTCCAATTAGGAATA
CCACACCCAGCAGGGTTAAAAAAGAAAAATCAGTGCAGTACTGGATGGGAGATGCATA
TTTTCAGTCCCTTATAGATGAGAGCTCAGAAAATATACTGCATTCAACCATACCTAGTATAAAC
AATGAAACACCCAGGGATTAGATATCAATATAATGTTCTCCACAGGGATGGAAGGATCACC
AGCAATATTCCAGAGTAGCATGACAAGAATCTTAGAGCCCTTAAAGAACACAAAACCCAGAAG
TAGTTATCTATCAATATATGGATGACTTATATGTAGGACTGACTTAGAAATAGGGCACATA
GAGAAAAATAGAGGAGTTAACAGGACACCTATTGAAATTGGGATTACACACCAAGACAAG
AAACATCAGAAAGAACCCCCATTCTTGGATGGGTATGAACTCCATCCTGACAAATGGACA
GTACAGCCTATAACAGCTGCCAGAAAAGGAGAGCTGGACTGTCAATGATATAACAGAAGTTAGT
GGGAAAGTTAAACTGGCAAGTCAGATTACCCAGGGATTAAAGTAAGGCAACTGTGAAAC
TCCTTAGGGGAGCCAAAGCACTAACAGACATAGTGCCTACTGACTGAAGAAGCAGAATTAGAA
TTGGCTGAGAACAGGGAAATTCTAAAAGAACAGTACATGGAGTATTATGACCCATCAAA
AGATTTAATAGCTGAAATACAGAACAGGGAAATGACCAATGGACATATCAAATTACCAAG
AACCATTTAAAATCTGAGAACAGGAAAGTATGCAAAATGAGGACTGCCACACTAATGAT
GTGAAACAGTTAGCAGAGGCAGTGCACAAAGATAACCCAGGGAAAGCATAGTAATATGGGGAA
AAACTCTAAATTAGACTACCCATCCAAAAGAAAATGGAGACATGGTGGTCAGACTATT
GGCAAGCCACCTGGATTCCCTGAGTGGAGTTGTCAATACCCCTCCCTAGTAAAATTGTGGT
ACCAAGCTGGAAAAGAACCCATAGTAGGGCAGAAACTTCTATGTAGATGGAGCAGCCAAT
AGGAAACTAAAATAGAAAAGCAGGGTATGCACTGACAAAGGAAGGAGCAGAAAGTGTITTC
CTTCACTGAAACAACAAATCAGAACAGACTGAATTACAAGCAATTAGCTAGCTTGCAGGATTC
AGGGCCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAAC
CAGATAAGAGTGAATCAGAATTAGTCAGTCAAATAATAGAACAGTTGATAAAAAGGAAAAAA
GTCTACCTATCATGGTACCAAGCACATAAAGGAATTGGAGGAATGAAACAAGTAGACAAATT
AGTAAGTAGTGGAAATCAGAAAAGTACTGTTCTAGATGGAATAGATAAAGCTCAAGAAGAGC
ATGAAAAATATCACAGCAATTGGAGAGCAATGGCTAGTGGATTTAATCTGCCACCCATAGTA
GCAAAGAAAATAGTAGCCAGCTGTGATAATGTCAGCTAAAGGGGAAGCCATGCATGGACA
AGTCGACTGTAGTCCAGGAATATGCCAATTAGACTGTACACATTAGAAGGAAAATCATCCT
AGTAGCAGTCATGTAGCCAGTGCTACATGGAAGCAGAGGTATCCAGCAGAAACAGGAC
AAGAAACAGCATACTTTAATCTAAAGTAGCAGGAAGATGGCAGTCAGTAAAGTAATACATACA
GATAATGGCAGTAATTCCACCACTGACAGCAGTAAAGGAGCAGCTGTTGGTGGGCAGATATCCAA
CGGAAATTGGAATTCCCTACAATCCCCAAAGTCAGGAGTAGTAGAATCCATGAATAAAGA
ATTAAGAAAATCATAGGGCAAGTAAGAGATCAAGCTGAGCACCCTAAGACAGCAGTACAAA
TGGCAGTATTCTCATTCAAAATTAAAAGAAAAGGGGGATTGGGGGTACAGTGCAGGGAG
AGAATAATAGACATAATAGCATCAGACATACAAACTAAAGAATTACAAAACAAATTATAAA
AATCAAAATTTCGGGTTTATIACAGAGACAGCAGAGACCCATTGGAAGAGGACCAGCCAA
ACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGTATATAAGGTAGTAC
CAAGAAGGAAAGCAAAAATCTTAAGGACTATGGAAAACAGATGGCAGGTGCTGATTGTGTG
GCAGGTAGACAGGATGAAGAT

FIGURE 75

RevExon1_TV2_C_ZAopt (SEQ ID NO:105)

ATGGCCGGCCGCAGCGGCGACAGCGACGAGGCCCTGCTGCAGGCCATCAAG
ATCATCAAGATCCTGTACCAGAGC

FIGURE 76

RevExon1_TV2_C_ZAwT (SEQ ID NO:106)

ATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGCTCCTCCAAGCAATAAAG
ATCATCAAGATCCTCTACCAAAGCA

FIGURE 77

751

SF162	(725)	EQDRDRSIRLAVLLEAIIWDDRSICVASSPPSPLEAAARIVPVER-	800
TV1.8_2	(743)	EQDRDRSIRLAVLLEAIIWDDRSICVASSPPSPLEAAARIVPVER-	
TV1.8_5	(745)	EQDRDRSIRLAVLLEAIIWDDRSICVASSPPSPLEAAARIVPVER-	
TV2.12-5/1	(730)	EQDSSEIIRLAVLLEAIIWDDRSICVASSPPSPLEAAARIVPVER-	
Consensus	(751)	EQDRDRSIRLVSGLSLAWDDLSLCLFSYHRLRDFILIAVRARELLGHS	

801

SF162	(774)	-RGWEAEKQWGNHICYVIQEDKNSAVSLFDAAIAAVARGTDRIIE	850
TV1.8_2	(793)	SLRGLQRGEWEILKYLGSIVQYVGLSHIKKSNAISLLDTIAIAVARGTDRIIE	
TV1.8_5	(795)	SLRGLQRGEWEILKYLGSIVQYVGLSHIKKSNAISPLTIAIAVARGTDRIIE	
TV2.12-5/1	(780)	SLRGLQRGWGTIEKYLGSIVQYVGLSHIKKSNAISLLDTIAIAVARGTDRIIE	
Consensus	(801)	SLRGLQRGEWEILKYLGSIVQYVGLSHIKKSNAISLLDTIAIAVAEGTDRIIE	

851

SF162	(818)	VADGRIGRAFLHIPPRIRQGEFAALL-	876
TV1.8_2	(843)	LVQRICRAILNIPPRIRQGEFAALL-	
TV1.8_5	(845)	LVQRICRAILNIPPRIRQGEFAALL-	
TV2.12-5/1	(830)	FIONTCRGIRNVPRIRQGEFAALO-	
Consensus	(851)	LVQRICRAILNIPPRIRQGEFAALL	

BEST AVAILABLE COPY

* 351 * * 400

SF162	(335)	KQLVTKLQAQFGNKEI-NVKQSSGGDPEEVMSHESNGGELEYCNSITQEN
TV1.8_2	(349)	QQVMKKIGEHRPNKII-IKFPHAGGDLEITMHSFNCRGEFFYCNTSNLFN
TV1.8_5	(349)	QQVMKKIGEHRPNKII-IKFPHAGGDLEITMHSFNCRGEFFYCNTSNLFN
TV2.12-5/1	(335)	QRVSOKIQELSPNSELKEAPHSGGDLEITTHSNGGEFFYCNTHDLFN
Consensus	(351)	QQVMKKLQEHEFPNKT IKFKPHAGGDLEITMHSFNCRGEFFYCNTSNLFN

401 * * ↓ β20/β21 ↓ 450

SF162	(384)	STWNN-----HIGPN-NNGTETIPGRHKOIGINRWOEVGKAMYAPPTRG
TV1.8_2	(398)	STYHS---NNGIJKYNGNSSSPITLLOCKIKOILVRMNOGVGOAMYAPPTRG
TV1.8_5	(398)	STYMP---KNGIJKYNGNSSSPITLLOCKIKOILVRMNOGVGOAMYAPPTRG
TV2.12-5/1	(385)	STYSNGTCTNGTCMSN--NTERITLLOCKIKOILNMGEVGRAMYAPPTRG
Consensus	(401)	STYHN NGTYKYNGNSS PITLOCKIKOILRMNOGVGQAMYAPPTRG

* 451 * * * 500

SF162	(427)	OIRGSSNEWGHIAHTRDGGKEISNT--TETFRPGGDMRDNWRSELYKYKV
TV1.8_2	(445)	NITCRSNITGILLTRDGGFNTNN--TETFRPGGDMRDNWRSELYKYKV
TV1.8_5	(445)	NITCRSNITGILLTRDGGFNTNN--TETFRPGGDMRDNWRSELYKYKV
TV2.12-5/1	(433)	NITCRSNITGILLTRDGGDNNTT--TETFRPGGDMRDNWRSELYKYKV
Consensus	(451)	NITCRSNITGILLTRDGGFNTNT TETFRPGGDMRDNWRSELYKYKV

501 550

SF162	(475)	VKEPLIGVAPTKAKRRVWOREKRAVITGAMELSGIAGGSTMGASITLT
TV1.8_2	(493)	VETKPLIGIAPTKAKRRVWOREKRAVIGIGAVFEGIAGGSTMGAASTITLT
TV1.8_5	(495)	VETKPLIGIAPTKAKRRVWOREKRAVIGIGAVFEGIAGGSTMGAASTITLT
TV2.12-5/1	(480)	VETKPLIGVAPTKAKRRVWOREKRAVIGIGAVFEGIAGGSTMGAASTITLT
Consensus	(501)	VEIKPLIGIAPTKAKRRVVWOREKRAVIGAVFLGFLGAAGSTMGAASITLT

551 600

SF162	(525)	VQAROHLSGIVOOCONNITRATTAQOHLQOLTVWGIKOLOARVLAERYLK
TV1.8_2	(543)	VQAROHLSGIVOOCONNITRATTAQOHLQOLTVWGIKOLOARVLAERYLK
TV1.8_5	(545)	VQAROHLSGIVOOCONNITRATTAQOHLQOLTVWGIKOLOARVLAERYLK
TV2.12-5/1	(530)	VQAROHLSGIVOOCONNITRATTAQOHLQOLTVWGIKOLOARVLAERYLK
Consensus	(551)	VQARQLLSGIVQQQSNNLKAQEAQHMLQLTWVGIKQLOARVLAERYLK

* 601 * * * 650

SF162	(575)	DQQLIGINGCSGKICLTTAVPNSSWSNKSLDQFWNDNMTWMOQWDREISNY
TV1.8_2	(593)	DQQLIGINGCSGKICLTTAVPNSSWSNASEKDFWDNMTWMOQWDREISNY
TV1.8_5	(595)	DQQLIGINGCSGKICLTTAVPNSSWSNASEADFWDNMTWMOQWDREISNY
TV2.12-5/1	(580)	DQQLIGIWCWSGKICLTTAVPNSSWSNKTOEDFWDNMTWMOQWDREISNY
Consensus	(601)	DQQLIGIWCWSGKICLTTAVPNSSWSNKSEADIWDNMTWMOQWDREISNY

651 700

SF162	(625)	TNTIYTTLIEESONQOEKNEQEELEKWAQWWWDISKNLWYIKIFIMI
TV1.8_2	(643)	TGLHTYNILEDSONQOEKNEKDIEELEKWNNEWWWDISNWPAWIKIFIMI
TV1.8_5	(645)	TNTIYTLEEDSONQOEKNEKDIEELEKWNNEWWWDISNWPAWIKIFIMI
TV2.12-5/1	(630)	TNTIYTLEEDSGQGERNEKDIAFWDNNEWWWEHINWYIKIFIMI
Consensus	(651)	TNTIYTLLEDSONQOEKNEKDLELDKWNNLNWFDISNWLYIKIFIMI

701 750

SF162	(675)	VGGIGLRLIIFAVLSIVNRVRQGYSPLSFQTLTPSPRGDRIGGIEEEGG
TV1.8_2	(693)	VGGIGLRLIIFAVLSIVNRVRQGYSPLSFQTLTPSPRGDRIGGIEEEGG
TV1.8_5	(695)	VGGIGLRLIIFAVLSIVNRVRQGYSPLSFQTLTPSPRGDRIGGIEEEGG
TV2.12-5/1	(680)	VGGIGLRLIIFAVLSIVNRVRQGYSPLSFQTLTPSPRGDRIGGIEEEGG
Consensus	(701)	VGGIGLRLIIFAVLSIVNRVRQGYSPLSFQTLTPSPRGDRIGGIEEEGG

BEST AVAILABLE COPY

↓: is the regions for β-sheet deletions

*: is the N-linked glycosylation sites for subtype C TV1 and TV2. Possible mutation (N→ Q) or deletions can be performed.

		1		50
SF162	(1)	---MDAMKRGLCCVLLLECGAVFVSPSAVKEWVTVAYGVDPWKEATPHL		
TV1.8_2	(1)	MRVMGTQKNCQQWWIWGILGFWMMLMICNTED	↓	WVTVAYGVDPWRAKUH
TV1.8_5	(1)	MRVMGTQKNCQQWWIWGILGFWMMLMICNTED	↓	WVTVAYGVDPWREAKUH
TV2.12-5/1	(1)	MRARGILKNYRHHWIWGILGFWMMLMMCNVKG	↓	WVTVAYGVDPVGREAKUH
Consensus	(1)	MRVMGTQKNCQQWWIWGILGFWMMLMICNVEDLWVTYGVDPWREAKUH		
		51	*	100
SF162	(47)	FCASDAKAYDTEVHNWATHACVPTDPNPQEIVLGNVTENFNMWKNNMVD		
TV1.8_2	(51)	FCASDAKAYMETEVHNWATHACVPTDPNPQEIVLGNVTENFNMWKNDMAD		
TV1.8_5	(51)	FCASDAKAYMETEVHNWATHACVPTDPNPQEIVLGNVTENFNMWKNNMAD		
TV2.12-5/1	(51)	FCASDAKAYEKEVHNWATHACVPTDPNPQEIVLGNVTENFNMWKNDMVD		
Consensus	(51)	FCASDAKAYETEVHNWATHACVPTDPNPQEIVLGNVTENFNMWKNNMVD		
		β2/V1V2/β3	*	
		101	↓	* * * * * 150
SF162	(97)	OMHEEDIISLWDOSIKPCKVLTPLCWTLNCTNTVGNRTVTGNSNSN		--SSN--
TV1.8_2	(101)	OMHEDVFSIWDOSIKPCKVLTPLCWTGNCDTNTVGNRTVTGNSNTNTNG		
TV1.8_5	(101)	OMHEDIISLWDOSIKPCKVLTPLCWTGNCDTNTVGNRTVTGNTNDTNIA		
TV2.12-5/1	(101)	OMOEDIISLWDOSIKPCKVLTPLCWTGNATVNYN		--NHS--
Consensus	(101)	QMHEDIISLWDQSLKPCPVLTPLCVTLNCTNTVGNRTVTGNSNSN A		
		151	* * * 200	
SF162	(139)	WKEMDRGEIKKNCSEKVKITISRNKMKSYALRYKIDIVPLDN		--DNTSE
TV1.8_2	(151)	TGIYNIEEMKNGSNAITELRDKKHKEYAEGSRDIVPLN		--ENSDNFTE
TV1.8_5	(151)	NATYKYEEMKNCSENAITELRDKKHKEYALRYKIDIVPLN		--ENSNNTFT
TV2.12-5/1	(141)	-----KDMKNCSEYVTHELRDKKKCNEALYRDLIVLNNRKNGNINN		
Consensus	(151)	A Y EEMKNCSEFNVTTEL RDKKHKEYALFYKLDIVPLNN ENSNNFTY		
		201	* ↓ * * * 250	
SF162	(185)	RLINCNTSTITQACPKVSFDPIPPIHYCAPAGYAI		LGKCNKT
TV1.8_2	(199)	RLINCNTSTITQACPKVSFDPIPPIHYCAPAGYAI		LGKCNKT
TV1.8_5	(199)	RLINCNTSTITQACPKVSFDPIPPIHYCAPADYAI		LGKCNKT
TV2.12-5/1	(185)	RLINCNTSAITQACPKVSFDPIPPIHYCAPAGYAI		LGKCNKT
Consensus	(201)	RLINCNTSTITQACPKVSFDPIPPIHYCAPAGYAI		LGKCNKT
		251	* * * 300	
SF162	(235)	VSTVOQCTHGLKPVVSTQLLLNGSLAEGIIIRSENLTENTKTIVHLNES		
TV1.8_2	(249)	VSTVOQCTHGLKPVVSTQLLLNGSLAEGIIIRSENLTENTKTIVHLNES		
TV1.8_5	(249)	VSTVOQCTHGLKPVVSTQLLLNGSLAEGIIIRSENLTENTKTIVHLNES		
TV2.12-5/1	(235)	VSTVOQCTHGLKPVVSTQLLLNGSLAEGIIIRSENLTENTKTIVHLNES		
Consensus	(251)	VSTVOQCTHGLKPVVSTQLLLNGSLAEGIIIRSENLTENTKTIVHLNES		
		301*	* * * 350	
SF162	(285)	VEINCTRPNNNTRKSVRIGPGQAFYATNDIIGNIRQAHCNISTDRWNKTL		
TV1.8_2	(299)	VEINCTRPNNNTRKSVRIGPGQAFYATNDIIGNIRQAHCNISTDRWNKTL		
TV1.8_5	(299)	VEINCTRPNNNTRKSVRIGPGQAFYATNDIIGNIRQAHCNISTDRWNKTL		
TV2.12-5/1	(285)	VEINCTRPNNNTRKSVRIGPGQAFYATNDIIGNIRQAHCNISTDRWNKTL		
Consensus	(301)	VEINCTRPNNNTRKSVRIGPGQAFYATNDIIGNIRQAHCNISTDRWNKTL		

BEST AVAILABLE COPY

NefD125G-Myr_TV2_C_ZAopt (SEQ ID NO:135)

ATGGCCGGCAAGTGGAGCAAGAGCAGCATCATCGGCTGGCCCGAGGTGC
GAGCGCATCCGCCGCACCGCAGGCCGCCAGGGCGTGGCAGGCCAGC
CAGGACCTGGAGAAGCACGGGCCCTGACCACCAGCAACACCGCCCACA
AACGCCGCTGCGCTGGCTGGAGGCCAGGAGGAGGAGGGCGAGGTGG
TTCCCCGTGCGCCCCCAGGTGCCCTGCCCTGACCTACAAGGCCGC
CGACCTGAGCTTCTCCTGAAGGAGAAAGGGCGGCCCTGGAGGGCCTG
AGCAAGAAGGCCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGG
TTCTTCCCCGGCTGGCAGAACTACACCCCCGGCCCCGGCGTGC
GACCTTCGGCTGGTACTTCAAGCTGGAGCCCCTGGACCCCCCG
GAGGCCAACGAGGGCGAGAACAACTGCCTGCTGCACCCATGAGCC
GGCATGGAGGACGAGGACCGCGAGGTGCTGCGCTGGAAGTT
CTGGCCCGCCACATGGCCCGAGCTGCACCCGAGTACTACAAGGACT
GCTGA

FIGURE 104

NefD125G_TV2_C_ZAopt (SEQ ID NO:134)

ATGGCGGCAAGTGGAGCAAGAGCAGCATCATGGCTGGCCCGAGGTGC
GAGCGCATCCGCCGCACCCGCAAGCGCCGCCAGGGCGTGGGCAGCGCCAGC
CAGGACCTGGAGAACGACGGGCCCTGACCACCAGCAACACCGCCCCACAAC
AACGCCGCTGCGCCTGGCTGGAGGCCAGGAGGAGGAGGGCGAGGTGGC
TTCCCGTGCCTGGCTGGAGGCCAGGTGGCTGGCCCTGCGCCCATGACCTACAAGGCC
CGACCTGAGCTTCTCCTGAAGGAGAACGGCGGCCCTGGAGGGCTGATCTAC
AGCAAGAACGCCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGC
TTCTTCCCCGGCTGGCAGAACTACACCCCCGGCCCGCGTGCCTCCCCCT
GACCTTCGGCTGGTACTCAAGCTGGAGCCCCTGGACCCCCCGAGGTGGAG
GAGGCCAACGAGGGCGAGAACAACTGCCTGCTGCACCCCATGAGCCAGCAC
GGCATGGAGGACGAGGACCGCGAGGTGCTGCCTGGAAGTTCGACAGCACC
CTGGCCCGCCGCCACATGGCCCGCGAGCTGCACCCGAGTACTACAAGGACT
GCTGA

FIGURE 103

gp140mod.TV1.wtLnative (SEQ ID NO:133)

1 gaattcatga gagtatggg gacacagaag aattgtcaac aatggtgat atggggcata
61 ttaggcitct ggatgctaattttgtaaaccggaggacc tgggggtgac cgftactac
121 ggctgtccccg tggtggcgca cggcaagacc accctgttct ggcggcagcga cggcaaggcc
181 tacgagaccg aggtgcacaa ctgtgggccc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgtggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gcccggcaca tgacggagga ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgacccgaca ccaacgtgac cggcaaccgc
421 accgtgaccc gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccaccacc gagctgccc acaagaagca caaggagttac
541 gcccctgttct accgcctgga catcggtccc ctgaacgaga acagcgacaa cttcacctac
601 cgcctgtatca actgaaacac cagcaccatc acccaggctt gcccccaaggt gagcttcgac
661 cccatccccca tccactactg cgcccccggcc ggctacgcca tccgtgaatg caacaacaag
721 accttcaacg gcacccggccc ctgttacaac gtgaggcccg tgcagtgcac ccacggcatc
781 aagccctgtgg tgagcacccca gctgtgtgt aacggcagcc tggccgagga gggcatcatc
841 atccgcagcg agaacctgac cgagaacacc aagaccatca tcgtgcacctt gaacgagagc
901 gtggagatca actgcaccccg ccccaacaac aacaccgcgca agagcgtgcg catcgcccc
961 ggccaggcccttacgcccac caacgacgtg atcggcaaca tccggccaggccactgcaac
1021 atcagcaccg accgctggaa caagaccctg cagcagggtg tgaagaagct gggcgagcac
1081 ttcccccaaca agaccatcca gttaagccc cacggccggcg ggcacctggaa gatcaccatg
1141 cacagttca actgcgcgg cgagttttt tactgcaaca ccagcaacctt gtcaacagc
1201 acctaccaca gcaacaacgg cacctacaag tacaacggca acagcagcag cccatcacc
1261 ctgcagtgcgca agatcaagca gatgtgcgc atgtggcagg gcfggccca gcccacccatc
1321 gccccccca tcggccggca catcacctgc cgcagcaaca tcaccggcat cctgtgtacc
1381 cgcacggcg gcttcaacac caccaacaac accgagacctt tccggccggcgccggccgac
1441 atgcgcgaca actggcgccag cgagctgtac aagtacaagg tggggat caagccccctg
1501 ggcacatggccccc ccaccaaggca aacgcggccgc gtgggtgcgc ggcggccggcc
1561 atcggccggcg tggccggggcttccctgggc gcccggccgc gacccatggg cgccggccagc
1621 atcaccctgaa cctgtgcaggcc cccggccatgt ctgagccggca tcgtgcagca gcaac
1681 ctgcgtgaagg ccatcgaggcc cccggccatgtc atgcgtgcgc tgaccgtgtt gggcatcaag
1741 cagctgcagg cccggccgtgtt ggcacatgcg cgcgttgcgtt aggaccagca gctgtggcc
1801 atctggggctt gcaacggccgc cctgtatgc accaccggccg tggccctggaa cagcagctgg
1861 agcaacaaga gcaacggccgtgtt ggcacatgc accaccggccg tggccctggaa cagcagctgg
1921 atcagcaactt acaccggccgtgtt gcaacggccgtgtt ggcacatgc accaccggccg
1981 aagaacgaga aggacccgtgtt ggagctggac aatggaaaca acctgtggaa ctgggttcgac
2041 atcagcaactt gggccctggta catctaactc gag

FIGURE 102

gp140mod.TV1 (SEQ ID NO:132)

1 gaattc~~tatgc~~ gcgtgatggg caccagaag aactgcc~~ac~~ agtgg~~tggat~~ ctggggcatc
61 ctgggcttct ggat~~gct~~gat gat~~c~~tcgaac accgaggacc t~~tg~~ggg~~tgac~~ cgtgtactac
121 ggc~~gt~~gccc~~g~~ tg~~tg~~g~~cg~~ca cgccaa~~gacc~~ acc~~c~~ttct gc~~cc~~ag~~cg~~ca cgccaa~~gg~~cc
181 tacgagaccg agg~~tg~~ca~~caa~~ c~~gt~~tg~~gg~~cc acc~~cac~~gc~~c~~ g~~cg~~tg~~cc~~ac cgac~~cc~~aa~~ac~~
241 ccccaggaga tc~~gt~~g~~ct~~ggg caac~~gt~~gacc gagaact~~ta~~ acat~~tg~~gaa ga~~ac~~gac~~at~~g
301 gccgacc~~g~~ga tg~~cac~~gagg~~a~~ c~~gt~~gat~~c~~ag~~c~~ ct~~gt~~gg~~acc~~ ag~~ag~~ct~~g~~aa~~g~~ cc~~ct~~g~~cg~~tg
361 a~~a~~g~~c~~t~~g~~ac~~cc~~ cc~~c~~t~~g~~tc~~g~~t g~~ac~~c~~c~~t~~g~~aa~~c~~ tg~~cac~~g~~g~~aca c~~ca~~a~~c~~gt~~g~~ac~~c~~ cg~~g~~ca~~ac~~cc~~g~~c
421 acc~~gt~~gac~~cc~~ g~~ca~~a~~c~~ac~~ac~~ caaca~~ac~~acc~~a~~ a~~a~~c~~g~~gc~~ac~~cg~~g~~ g~~ca~~t~~ct~~aca~~a~~ cat~~cg~~agg~~g~~ag
481 at~~g~~a~~a~~g~~a~~act~~g~~c~~at~~tc~~a~~ cgcc~~ac~~cc~~ac~~ g~~ag~~ct~~g~~cg~~cg~~ a~~c~~a~~a~~g~~a~~g~~c~~a ca~~agg~~g~~ag~~tg~~a~~c
541 g~~cc~~c~~t~~gt~~t~~t~~c~~ acc~~g~~c~~c~~t~~g~~ga~~c~~ cat~~g~~t~~g~~cc~~c~~ ct~~g~~a~~ac~~g~~g~~aca~~a~~ c~~tc~~cac~~c~~t~~ac~~
601 c~~gc~~c~~t~~gat~~ca~~ a~~t~~g~~c~~ca~~ac~~ac~~c~~ c~~ag~~c~~ac~~cat~~c~~ acc~~c~~agg~~cc~~ g~~cc~~ca~~ag~~gt~~g~~ g~~ag~~ct~~tc~~g~~ac~~
661 cccat~~cccc~~ tccact~~act~~tg~~g~~ c~~g~~ccccc~~g~~cc gg~~ct~~ac~~g~~cc~~a~~ t~~c~~ct~~g~~a~~ag~~tg~~c~~ caaca~~ac~~aa~~g~~
721 ac~~c~~t~~t~~ca~~ac~~g~~c~~ g~~c~~acc~~gg~~cc~~c~~ ct~~g~~ct~~a~~ac~~a~~ g~~t~~g~~ag~~cc~~ac~~g~~c~~ t~~g~~c~~ag~~t~~g~~ac~~c~~ cc~~ac~~gg~~cc~~at~~c~~
781 a~~a~~g~~cc~~gt~~gg~~ tg~~ag~~g~~ac~~cc~~a~~ g~~c~~t~~g~~ot~~g~~ct~~g~~ a~~a~~c~~g~~g~~c~~ag~~c~~ t~~g~~g~~cc~~g~~ag~~ga~~g~~ g~~gg~~cat~~c~~at~~c~~
841 at~~cc~~gc~~ag~~cg~~g~~ ag~~a~~ac~~c~~tg~~ac~~ cg~~ag~~aa~~c~~acc~~a~~ a~~a~~g~~ac~~cat~~a~~ t~~c~~tg~~g~~ac~~ct~~ g~~a~~c~~g~~g~~ag~~ag~~c~~
901 gt~~gg~~g~~ag~~at~~ca~~ a~~t~~g~~c~~ac~~cc~~g~~c~~ c~~cc~~ca~~a~~ac~~a~~ a~~a~~c~~ac~~cc~~g~~ca~~a~~ ag~~ag~~ct~~g~~cg~~g~~ c~~at~~gg~~cc~~cc
961 gg~~cc~~agg~~cc~~ t~~c~~t~~ac~~g~~cc~~ac~~a~~ ca~~a~~c~~g~~ac~~tg~~ t~~at~~cg~~g~~ca~~a~~ t~~c~~cc~~g~~cc~~ag~~g~~c~~ cc~~act~~g~~ca~~ac~~a~~
1021 at~~c~~ag~~c~~ac~~cc~~g~~c~~ acc~~g~~ct~~gg~~aa~~c~~ ca~~a~~g~~ac~~c~~ct~~g~~c~~ c~~ag~~c~~ag~~gt~~g~~ t~~g~~a~~ag~~g~~a~~g~~c~~t g~~gg~~g~~ag~~g~~ac~~
1081 tt~~cc~~ca~~a~~ca~~a~~ ag~~ac~~cat~~c~~ca~~a~~ g~~t~~ca~~ag~~cc~~c~~ c~~ac~~g~~cc~~gg~~g~~ g~~g~~g~~ac~~ct~~gg~~ga~~a~~ g~~a~~t~~c~~ac~~ca~~at~~g~~
1141 cac~~ag~~ct~~ca~~ a~~t~~g~~cc~~g~~cg~~g~~c~~ cg~~ag~~tt~~tc~~ t~~a~~c~~g~~ca~~a~~cc~~a~~ cc~~ag~~ca~~ac~~ct~~g~~ t~~g~~ta~~ca~~ac~~ag~~g~~c~~
1201 ac~~c~~t~~t~~acc~~c~~aca~~a~~ac~~gg~~g~~c~~ cac~~c~~t~~a~~ca~~a~~g~~t~~ t~~a~~c~~a~~ac~~gg~~ca~~a~~ ac~~a~~g~~c~~ag~~cg~~g~~c~~ c~~cc~~cat~~c~~acc~~a~~
1261 ct~~g~~c~~ag~~tg~~ca~~ ag~~a~~at~~ca~~ag~~g~~ca~~a~~ g~~a~~t~~c~~g~~tg~~g~~cg~~gc~~a~~ at~~tg~~g~~gc~~agg~~a~~ g~~g~~cc~~ac~~cc~~t~~ac~~a~~
1321 g~~cc~~cc~~cc~~cc~~cc~~ t~~c~~g~~cc~~gg~~cc~~aa~~c~~ cat~~c~~ac~~c~~t~~g~~ c~~g~~c~~ag~~ca~~a~~ca~~a~~ t~~c~~ac~~cc~~gg~~cc~~at~~c~~ c~~ct~~g~~ct~~g~~ac~~
1381 c~~g~~gc~~ac~~gg~~cc~~g~~c~~ g~~c~~t~~ca~~ac~~ac~~ ca~~a~~cca~~ac~~ac~~a~~ acc~~g~~ag~~ac~~ct~~g~~ t~~c~~cc~~cc~~gg~~cc~~ c~~gg~~cc~~gg~~g~~c~~ac~~a~~
1441 at~~g~~cg~~cg~~aca~~a~~ act~~gg~~g~~cg~~ag~~c~~ cg~~ag~~et~~g~~ta~~c~~ a~~a~~g~~t~~aca~~gg~~ t~~g~~g~~tg~~g~~ag~~at~~c~~ ca~~ag~~cc~~cc~~ct~~g~~
1501 gg~~cat~~g~~cc~~cc~~c~~ ca~~ac~~ca~~a~~agg~~c~~ ca~~a~~ag~~cg~~cc~~gc~~ g~~t~~g~~gt~~g~~cg~~ac~~c~~ g~~c~~g~~ag~~aa~~g~~cg~~c~~ cg~~cc~~gt~~gg~~gc
1561 at~~c~~cg~~cg~~cc~~c~~ tg~~tt~~cc~~ct~~gg~~gg~~ ct~~tc~~ct~~gg~~gg~~c~~ g~~cc~~cc~~gg~~gg~~c~~ g~~c~~ac~~cc~~at~~gg~~ g~~g~~cc~~gg~~cc~~c~~ag~~c~~
1621 at~~c~~ac~~cc~~ct~~g~~a~~a~~ c~~cg~~tg~~ca~~gg~~g~~ c~~cg~~cc~~ag~~ct~~g~~ ct~~g~~ag~~cg~~gg~~c~~ t~~c~~tg~~g~~cg~~ac~~g~~c~~ g~~c~~ag~~ag~~ca~~a~~
1681 ct~~g~~ct~~g~~g~~ag~~g~~c~~ cc~~at~~cg~~ag~~gc~~c~~ cc~~ag~~c~~ag~~ca~~a~~ at~~g~~ct~~g~~cg~~ac~~ g~~t~~ac~~cg~~tg~~tg~~ g~~g~~g~~ca~~t~~ca~~ag~~c~~
1741 c~~ag~~ct~~g~~cg~~ag~~g~~c~~ c~~cc~~cg~~tg~~gt~~c~~ g~~g~~cc~~at~~cg~~g~~ag~~c~~ cg~~ct~~ac~~ct~~g~~a~~ ag~~g~~acc~~ag~~ca~~a~~ g~~c~~tg~~ct~~gg~~g~~
1801 at~~c~~t~~g~~gg~~gg~~ct~~c~~ g~~ca~~g~~cg~~gg~~cc~~g~~c~~ c~~c~~t~~g~~at~~ct~~g~~c~~ acc~~ac~~cc~~g~~cc~~c~~ t~~g~~cc~~ct~~gg~~aa~~ c~~a~~g~~c~~ag~~ct~~g~~g~~
1861 ag~~ca~~aca~~a~~ga~~a~~ g~~g~~g~~ag~~aa~~g~~ga~~a~~ ca~~a~~cat~~gg~~g~~ac~~ a~~a~~cat~~g~~ac~~ct~~ g~~g~~at~~g~~cg~~at~~g~~c~~ gg~~ac~~cc~~g~~cg~~ag~~
1921 at~~c~~ag~~ca~~act~~a~~ ac~~a~~c~~cc~~gg~~cc~~ct~~c~~ g~~a~~t~~c~~at~~ca~~ac~~a~~ ct~~g~~tg~~gg~~agg~~a~~ ac~~a~~g~~c~~cc~~ag~~g~~g~~ a~~c~~cc~~g~~agg~~g~~ag~~c~~
1981 a~~a~~g~~a~~ac~~g~~g~~a~~ a~~g~~g~~a~~c~~t~~g~~c~~ g~~g~~ag~~ct~~gg~~ac~~ a~~a~~g~~t~~gg~~aa~~ a~~c~~c~~t~~g~~tg~~g~~aa~~ ct~~g~~g~~tt~~g~~ac~~
2041 at~~c~~ag~~ca~~act~~a~~ gg~~cc~~ct~~gg~~ta~~a~~ cat~~ct~~a~~act~~c~~g~~ag~~c~~

FIGURE 101

gp140mod.TV1.tpa1 (SEQ ID NO:131)

1 atggatgcaa tgaagagagg gctctgctgt gtgctgctgc tgggtggagc agtctcggt
61 tcgcccagcg ccagcacccga ggacctgtgg gtgaccgtgt actacggcggt gcccgtgtgg
121 cgcgacgcca agaccacccct gttctgcgcc agcgacgcca aggccatcgaa gaccgaggtg
181 cacaacgtgt gggccaccca cgcctgcgtg cccaccgacc ccaacccca ggagatcg
241 ctgggcaacg tgaccgagaa ctcaacatcg tggagaagacg acatggccga ccagatgcac
301 gaggacgtga tcagcctgtg ggaccagagc ctgaaggccct gctgtaaagct gaccccccgt
361 tgcgtgaccc tgaactgcac cgacaccaac gtgaccggca accgcaccgt gaccggcaac
421 agcaccaaca acaccaacgg caccggcatc tacaacatcg aggagatgaa gaactgcac
481 ttcaacgcca ccaccgagct gcgcgacaag aagcacaagg agtacgcct gttctacgc
541 ctggacatcg tgccctgaa cgagaacagc gacaacttc cctaccgcct gatcaactgc
601 aacaccagca ccatcacccca ggcctgcccc aaggtgagct tgacccat cccatccac
661 tactgcgccc cgcgggcta cgcgcattcg aagtgcacaaca acaagaccc caacggcacc
721 ggcccctgt acaacgttag caccgtcgtag tgacccacg gcatcaagcc cgtggtagc
781 acccagctgc tgctgaacgg cagcctggcc gaggaggca tcatcatcg cagcgagaac
841 ctgaccgaga acaccaagac catcatcgta caccgtgaaacg agagcgtgga gatcaactgc
901 acccgccccca acaacaacac cgcgaagago gtgcgcacgc gccccggccca ggccttctac
961 gccaccaacg acgtgatgg caacatccgc caggccccact gcaacatcg caccgaccgc
1021 tggaaacaaga cctgcagca ggtgatgaaag aagctggcg agcactccc caacaagacc
1081 atccagttca agccccacgc cggggcgac ctggagatca ccatgcacag ctcaactgc
1141 cgcggcgagt tcttctactg caacaccaggc aaccgttca acagcaccta ccacagcaac
1201 aacggcacct acaagttacaa cggcaacagc agcagccccca tcaccctgca gtgcaagatc
1261 aaggcagatcg tgccatgtg gcagggcgtag ggcgcaggccca cctacgcccc cccatcgcc
1321 ggcaacatca cctgcgcgcg caacatcacc ggcacatctgc tgacccgcga cggggcttc
1381 aacaccacca acaacaccga gacccgcgc cccggggcgcc ggcacatcg cgacaactgg
1441 cgcagcgagc tgatacaagta caaggtggtag gagatcaagc cccgtggcat cgcggccacc
1501 aaggcacaagc gcccgtggc gcagcgccg aagcgcccg tgggcacatcg cgcgtgttc
1561 ctgggcttcc tggggccgcgc cggcagcacc atggggccgcg ccagcatcac cctgaccgt
1621 caggccgcgc agtgcgttag cggcatcgta cagcagcaga gcaacctgt gaaaggccatc
1681 gaggccccgcg agcacatgt gcacgtgacc gtgtggggca tcaagcagct gcaggccccgc
1741 gtgcgtggca tcgagcgcta cctgaaggac cagcagctgc tgggcacatcg gggctgcagc
1801 ggccgcctga tctgcaccac cgcctgcgc tggaaacagca gctggagccaa caagagcg
1861 aaggacatcg gggacaacat gaccgtggatc cagtgccggacc ggcagatcg caactacacc
1921 ggcctgtatc acaacactgt ggaggacagc cagaaccaggc aggagaagaa cgagaaggac
1981 ctgcgtggagc tggacaagtg gaacaacctg tggaaactggc tgcacatcg caactggcc
2041 tggtatcatcg aa

FIGURE 100

FIGURE 99

1 gaattcatga gagtgatggg gacacagaag aattgtcaac aatggtggat atggggcatc
61 ttaggcttctt ggtgtcaat gatttgcatac acccgaggacc tggggfjac cgtgtactac
121 ggcgtgcccg tggcgccgca cgccaaagacc accctgtct gcgcgcgcga cgccaaaggcc
181 tacgagaccg aggtgcacaa cgtgtggcc acccacgcct gcgtgcccac cgaccccaac
241 cccccaggaga tcgtgtggg caacgtgacc gagaacitca acatgtggaa gaacgacatg
301 gccgaccaga tgacacgagga cgtgtacgc ctgtggacc agagcctgaa gcccgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcacccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatacaca catcgaggag
481 atgaagaact gcagttcaa cgccaccacc gagctgcgcg acaagaagca caaggagttac
541 gccctgtctt accgcctgga catcgccctt ctgaacgaga acagcgacaa ctgcacccatc
601 cgcctgtatca actgcaacac cagcaccatc acccaggccct gccccaaagggt gagctgcac
661 cccatccccca tccactactg cgccccccgc ggctacccca tcctgaatgt caacaacaag
721 acctcaacg gcacccggcc tcgtacaaac gtgaggaccc tgcaatgcac ccacggcatac
781 aagcccggtg tgaggaccca gtcgtgtc aacggcagcc tggccgagga gggcatcatc
841 atccgcagcg agaacctgac cgagaacacc aagaccatca tcgtgcacctt gaacgagagc
901 gtggagatca actgcacccg ccccaacaac aacacccgca agagcgtgcg catcgccccc
961 ggccaggccctt acacgcccac caacgacgtg atcggcaaca tccggcaggc ccactgcaac
1021 atcagcaccg accgctggaa caagaccctg cagcaggta tgaagaagct gggcgagcac
1081 ttccccaaaca agaccatcca gttcaagccc cacgcggcg ggcacctgga gtcaccatg
1141 cacagttca actgcgcgg cgagtcttc tactgcaaca ccagcaaccc ttcaacagc
1201 acctaccaca gcaacaacgg cacctacaag tacaacggca acagcagcag ccccatcacc
1261 ctgcagtgcata agatcaagca gatgtgcgc atgtggcagg gctgtggccca ggcacccatc
1321 gccccccca tcggccgaa catcacctgc cgcaccaaca tcaccggcat cctgctgacc
1381 cgcgacggcg gtcacacac caccaacaac accgagaccc tccggccggc cggcgccgac
1441 atgcgcgaca actggcgcag cgagctgtac aagtacaagg tggggagat caagccctg
1501 ggcacccccc ccaccaacgg caaggccgc gtggcgcgc gcgagaagcg cggcgccgg
1561 atcggcgcgg tggctggg cttccctggc gcccggca gcaccatggg cggccgcgc
1621 atcaccctga ccgtgcaggc cgcgcacgtc ctgacggcga tcgtgcagca gcagagcaac
1681 ctgcgtaaagg ccatcgaggc ccagcagcac atgctgcgc tgaccgtgt gggcatcaag
1741 cagctgcagg cccgcgtgtc ggccatcgag cgatccatgaa aggaccagca gctgtggc
1801 atctgggggt gcagcggccg cctgatctgc accacccgcg tggctggaa cagcagctgg
1861 agcaacaaga gcgagaagga catctggac aacatgaccc ttcatcatca tcgtggccgg cctgatcg
1921 atcagcaact acacccggccctt gatctacaac ctgctggagg acagccagaa ccagcaggag
1981 aagaacgaga aggacccgtt ggagctggac aagtggaaaca acctgtggaa ctggttcgc
2041 atcagcaact ggccctggta catcaagatc ttcatcatca tcgtggccgg cctgatcg
2101 ctgcgcatac tcttcgcgt gctgagcatc gtgacccgcg tggccaggcc ttcaccc
2161 ctgagcttcc agaccctgac ccccaaccc cgcggcctgg accgcctggg cggcatcg
2221 gaggagggcg gcgagcaggaa cgcgcaccgc agcatccgc tggcgccgg cttccctgac
2281 ctggcctggg acgacatgcgc caacccgtc ctgttcagctt accacccgcg cggcgacttc
2341 atcctgatcg ccgtgcgcgc cgtggagctt ctggccaca gcagcctgcg cggccctgc
2401 cgcggctggg agatctgaa gtacccggc accctggcgc agtactgggg cctggagctg
2461 aagaagagcg ccatcagccctt gctggacacc atcgcctatca ccgtggccga gggcaccgac
2521 cgcacatcatcg agtggcgcgc ggcacatcg cggccatcc tgaacatccc cgcgcgc
2581 cgccaggccctt tcgtgttccatc ctgcag

FIGURE 98

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtgat ctggggcatc
61 ctgggcttct ggatgtat gatctcaac accgaggacc tgggggtgac cgtgtactac
121 ggcgtcccg tggcgccga cgccaagacc accctgttct ggcgcagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtggcc acccacgcct gcgtgcccac cgacccaaac
241 cccccaggaga tcgtgtggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgtacgc ctgtgggacc agagcctgaa gccctgcgt
361 aagctgaccc ccctgtgcgt gaccctgaac tgacccgaca ccaacgtgac cggcaaccgc
421 accgtgaccc gcaacagcac caacaacacc aacggcaccc gcatctacaa catcgaggag
481 atgaagaact gcagctcaa cgccaccacc gagctgcgcg acaagaagca caaggagtac
541 gcccctgttct accgcctgga catgtgcct ctgaaacgaga acagcgacaa ctccacccatc
601 cgcctgtatca actgcaacac cagcaccatc acccaggccct gccccaaagggt gagcttcgac
661 cccatccccca tccactactg cgcgcgcgc ggcctacgcca tcctgaagtg caacaacaag
721 accttcaacg gcacccggccc ctgttacaac gtgagcaccg tgtagtgcac ccacggcatc
781 aagccctgtgg tgaggaccca gctgtgtcg aacggcagcc tggccgagga gggcatcatc
841 atccgcagcg agaacctgac cgagaacacc aagaccatca tcgtgcacctt gaacgagagc
901 gtggagatca actgcacccg ccccaacaac aacacccgcg agagcgtgcg catcgcccc
961 ggccaggccct tctacccac caacgacgtg atcggcaaca tccggcaggc ccactgcaac
1021 atcagcaccg accgctggaa caagaccctg cagcaggta tgaagaagct gggcgagcac
1081 ttcccaaca agaccatcca ttcaagccc cacgcccggcg ggcacccatgg gatcaccatg
1141 cacagcttca actggcgccg cgagttctc tactgcaaca ccagcaacctt gttcaacagc
1201 acctaccaca gcaacaacgg cacctacaag tacaacggca acagcagcag ccccatcacc
1261 ctgcagtgca agataaagca gatgtgcgc atgtggcagg gctggccca ggccacccatc
1321 gcccccccca tcggcgcaaa catcacctgc cgcagcaaca tcaccggcat cttgtgcacc
1381 cgcgcacggcg gttcaacac caccaacaac accgagaccc tccggccgg cggccggcgc
1441 atgcgcgaca actggcgacg cgagctgtac aagtacaagg tggtgagat caagccctg
1501 ggcacccccc ccaccaaggc caagcgccgc gtggcgacg ggcggaaaggcg cggcgccgg
1561 atcggcgccg ttttcttggg cttcttggc ggcgcggca gcaccatggg cggccggc
1621 atcaccctga cctgtcgaggc cgcgcacgtg ctgagcgca tcgtgcagca gcagagcaac
1681 ctgctgaagg ccatcgaggc ccagcagcac atgctgcagc tgaccgtgt gggcatcaag
1741 cagctgcagg cccgcgtgt ggcgcacgtg cgctacccatgaa aggaccagca gctgtgggc
1801 atctggggct gcagcgccg cctgtatgc accaccgcg tgccctggaa cagcagctgg
1861 agcaacaaga ggcggaaaggcatctggac aacatgaccc tggatgcgtg ggaccgcgag
1921 atcagcaact acaccggccct gatctacaac ctgtgtgggg acagccagaa ccagcaggag
1981 aagaacgaga aggacccgtg gtagtggacaa aagtggaaaca acctgtggaa ctgggtcgac
2041 atcagcaact ggcctggta catcaagatc ttcatcatga tcgtggccg cctgtatcgcc
2101 ctgcgcacca tcttcgcgt gtcgtatgc gtcgtatgc gtcgtatgc gtcgtatgc
2161 ctgagttcc agaccctgac ccccaacccgc cgcggcctgg accgcctgg cggcatcgag
2221 gaggaggccg ggcgcacggc cgcgcacccgc agcatccgc tggatgcgtg cttccatgc
2281 ctggccctggg acgacccgtg ccatccgtgc ctgttgcgtt accaccgcct ggcgcaccc
2341 atcctgtatgc cctgtatgc cgtggatgtg ctggccaca gtcgtatgc gtcgtatgc
2401 cgcggctggg agatccatgaa gtcgtatgc gtcgtatgc gtcgtatgc gtcgtatgc
2461 aagaagagcg ccatcgaccct gtcgtatgc atcggccatca ctcgtatgc gtcgtatgc
2521 cgcacatcg agtgggtgca ggcgcacccatcc tggatgcgtg cttccatgc
2581 cgcacatcg agtgggtgca ggcgcacccatcc tggatgcgtg cttccatgc

FIGURE 97

gp160mod.TV1.tpa1 (SEQ ID NO:125)

1 gtcgacgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgtct gctgtgtgga
61 gcagtctcg ttgcgccag cgccagcacc gaggaccgt ggggtgaccgt gtactacggc
121 gtgcggctgt ggcgcgacgc caagaccacc ctgttctgctg ccagcgacgc caaggcctac
181 gagaccgagg tgcacaacgt gtggccacc cacgcctgctg tgcccaccga ccccaacccc
241 caggagatcg tgcgtggcaaa cgtgaccgag aacttcaaca tgtgaagaa cgacatggcc
301 gaccagatgc acgaggacgt gatcagccctg tgggaccaga gcctgaagcc ctgcgtgaag
361 ctgacccccc tgcgtgtac cctgactgc accgacacca acgtgaccgg caaccgcacc
421 gtgacccggca acagcacccaa caacaccaac ggcacccggca tctacaacat cgaggagatg
481 aagaactgca gcttcaacgc caccaccggag ctgcgcgaca agaaggaccaa ggagtacgcc
541 ctgttctacc gcctggacat cgtgcctgt aacgagaaca gcgcacaactt cacttaccgc
601 ctgatcaact gcaacaccag caccatcacc caggcctgct ccaaggttag ctgcacccccc
661 atccccatcc actactgcgc ccccgccggc tacgcacatcc tgaagtgc aaacaagacc
721 ttcaacggca cggcccccctg ctacaacgtg agcaccgtgc agtgcacccca cggcatcaag
781 cccgtggta gcacccagct gctgtgaaac ggcagccctgg ccgaggaggg catcatcatc
841 cgcagcgaga acctgaccga gaacaccaag accatcatcg tgcacctgaa cgagagcgtg
901 gagatcaact gcacccgccc caacaacaac accccgaaga gcgtgcgcatt cggcccccggc
961 caggccttct acgcccacca cgcgtgatc ggcaacatcc gccaggccca ctgcaacatc
1021 agcaccgacc gctggaacaa gaccctgcag caggtagtga agaagctggg cgagcacttc
1081 cccaaacaaga ccatccaggta caagcccac gcccggcg acctggagat caccatgcac
1141 agcttcaact gccgcggcga gtttcttac tgcaacacca gcaacctgtt caacagcacc
1201 taccacagca acaacggcac ctacaagtg aacggcaaca gcagcagccccc catcaccctg
1261 cagtgcaga tcaaggcagat cgtgcgcatt tgccaggcg tgccaggcacc cacctacgc
1321 ccccccattcg cggcaacat cacccgcgc agcaacatca cggcatct gctgaccggc
1381 gacggcggct tcaacaccac caacaacacc gagaccccttgc ccccccggc cggcgcacatg
1441 cggcacaact ggcgcagcg agtgtacaag tacaaggtagg tgtagatcaa gccccctggc
1501 atgcggccca ccaaggccaa gcgcgcgtg gtgcagcgcg agaagcgccgc cgtggcattc
1561 ggcgcgtgt tcctggcctt cctggccgc gcccggcagca ccatggccgc cggcgcacatc
1621 accctgaccg tgcaggcccg ccagctgtg agccgcattcg tgcagcagca gagcaacatc
1681 ctgaaggcca tgcaggccca gcagcacatcg ctgcagctga ccgtgtggg catcaagcag
1741 ctgcaggccc gctgtgtgc catcgagcgc tacctgaagg accagcagct gctggcattc
1801 tggggctgca gcccggccct gatctgcacc accgcgtgc cctggaaacag cagctggagc
1861 aacaagagcg agaaggacat ctgggacaac atgacctgta tgcagtggg cccgcgagatc
1921 agcaactaca cccgcctgtat ctacaacccctg ctggaggaca gccagaacca gcaggagaag
1981 aacgagaagg acctgtgta gctggacaag tggaaacaacc tggaaactg gttcgacatc
2041 agcaactggc cctggatcat caagatctt atcatgatcg tggggccct gatcgccctg
2101 cgcacatct tcgcgtgt ggcacatcg aaccgcgtgc gcccggcta cagcccccctg
2161 agcttccaga ccctgacccca cagcccccgc ggcctggacc gcctggccgg catcgaggag
2221 gagggcggcg agcaggaccg cgaccgcacc atccgcctgg tgagcggctt cctgagcctg
2281 gcctggacg acctgcgcacca cctgtgcctg ttgcgttacc accgcgtgc cgcacttcatt
2341 ctgatcgccg tgcgcgcgt ggagctgtg ggcacacgca gcctgcgcgg cctgcagcg
2401 ggctggaga tccgtggatca cctggccgcg ctggcgtact gatctggccct ggagctgaaag
2461 aagagcgcca tcagccctgt ggacaccatc gccatcaccg tggccggaggg caccgaccgc
2521 atcatcgacc tggcgtggcgtg catctggccgc gccatcctgaa acatccccccg cgcacatccgc
2581 caggcgttcg aggccgcctt gctgtactc gag

FIGURE 96

gp160mod.TV1.mut7.delV2 (SEQ ID NO:124)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtgggtggat ctggggcatc
61 ctggcttct ggtatcgat gatctgcaac accgaggacc tgggggtgac cgtgtactac
121 ggcgtgcccc tggggcgca cgccaagacc accctgttct ggcggcagcga cgccaaggcc
181 tacggagaccg aggtgcacaa cgtgtggcc acccacgcct gcgtgcccac cgaccccaac
241 cccccaggaga tcgtgtggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgacgagga cgtgatcagc ctgtgggacc agagctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc
541 atcacccagg cctgccccaa ggtgagcttc gacccatcc ccatccacta ctgcgcccc
601 gccggctacg ccatctgaa gtgcaacaac aagaccctca acggcaccgg cccctgctac
661 aacgtgagca cctgtcagtg caccacggc atcaagcccg tggtagcac ccagctgtg
721 ctgaacggca gcctggccga ggagggcatc atcatccga cgcgagaacct gaccgagaac
781 accaagacca tcatctgca cctgaacgag agcgtggaga tcaactgcac cggccccaac
841 aacaacaccc gcaagagcgt ggcacatggc cccggccagg ctttacgc caccacgac
901 gtgatcggca acatccgcca ggcccactgc aacatcagca ccgaccgctg gaacaagacc
961 ctgcagcagg ttagtgaagaa gctggcgag cacttccca acaagaccat ccagtcaag
1021 cccacgcgcg gggcgaccc ggagatcacc atgcacagct tcaactgcgcg cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagcccccattc accctgcagt gcaagatcaa gcagatctg
1201 cgcacatgtggc agggcgtggg ccaggccacc tacgcggggg ccatcgccgg caacatcacc
1261 tgcgcgcagca acatcaccgg catctgtg acccgcgacg gggcttcaa caccaccaac
1321 aacaccgaga cttccgccc cggccggccg gacatcgccg acaactggcg cagcgagctg
1381 tacaagtaca aggtggtgg aatcaagccc ctggccatcg cccccaccaa ggcacatcagc
1441 agcgtggtgc agagcgagaa gagcgcgcgt ggcacatggc cctgtttctt gggcttctg
1501 ggcgcgcgcg gcagccat gggccgcgc accatcacc tacgcgtca gggccgcgc
1561 ctgcgtgagcg gcatctgca gcagcagcgc aacatgtga aggccatcga gggccagcag
1621 cacatctgcg acgtgaccgt gtggggcatc aagcagctgc aggccgcgt gctggccatc
1681 gagcgcgtacc tgaaggacca gcacatgtgg ggcacatggc gctgcacgcgg cccctgtatc
1741 tgcaccaccg ccgtgcccctg gaacagcagc tggagcaaca agagcgagaa ggacatctgg
1801 gacaacatga cttggatgca gtgggaccgc gagatcagca actacaccgg cttgatctac
1861 aacatctgtgg aggacagcca gaaccacgcg gagaagaacg agaaggaccc gctggagctg
1921 gacaagtgg acaacatgtg gaactggtc gacatcagca actggccctg gtacatcaag
1981 atcttcatca tgatctgtgg cggccatgc ggcacatcgca tcatcttcgc cgtgtgagc
2041 atcgtgaacc gcgtgcgcgc gggctacacgc cccctgatct tccagaccct gaccccccagc
2101 ccccgccgc tggaccgcct gggccgcac gaggaggagg gggccgcgc gggccgcac
2161 cgcacatcc gcctggtag cggccatgc ttcatctgca tcgcgtgcg cccctggag
2221 tgcctgttca gtcaccaccg cttgcgcgcac ttcatctgca tcgcgtgcg cccctggag
2281 ctgcgtggcc acacgcgcct ggcggccctg cggccggcgt gggagatccctt gaaatcttg
2341 ggcacatgtgg tgcgtactg gggccgtgg gtcgtggcc tcaagaaga ggcacatcg cctgtggac
2401 accatcgcca tcaccgtggc cgagggcacc gaccgcata tcgtgttgc gacgcgcac
2461 tgcgcgcgc tccatgtggccat ccccgccgc atccggcagg gcttcggagg cccctgttg
2521 taactcgag

FIGURE 95

gp160mod.TV1.delV2 (SEQ ID NO:123)

1 gaattcatgc gcgtgatggg caccagaag aactgccagc agtggatggat ctggggcata
61 ctgggccttc ggatgcttat gatctcaac accgaggacc tgggggtgac cgtfactac
121 ggcgtgcggc tggcgccga cgccaagacc accctgtttc gcccagcga cgccaaggcc
181 tacgagaccc aggtgcacaa cgtgtggcc acccagccct gctgtccac cgaccccaac
241 cccccaggaga tcgtgtggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtggacc agagcctgaa gcccgcgt
361 aagctgaccc ccctgtgcgt gacctgaaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccc gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagctcaa cggccggcc gggccctga tcaactgcaa caccagcacc
541 atcaccagg cctgccccaa ggtgagcttc gacccatcc ccatccacta ctgcgc
601 gcccgcatacg ccatctgaa gtgcaacaac aagacccatca acggcaccgg cccctgtac
661 aacgtgagca cctgtcactg caccacggc atcaagcccg tggtgaggac ccagctgt
721 ctgaacggca gcctggccga ggagggcatac atcatccgca gcgagaactt gaccgagaac
781 accaagacca tcatctgca cctgaacggag agcgtggaga tcaactgcac cggccccaac
841 aacaacaccc gcaagagcgt ggcacccggc cccggccagg ccttctacgc cacaacgac
901 gtgatcgcca acaitccgcca gcccactgc aacatcagca cgcaccgtg gaacaagacc
961 ctgcagcagg ttagtaagaa gctggccgag cacttccca acaagaccat ccagttcaag
1021 cccccacggc gggcgaccc ggagatcacc atgcacagct tcaactgcgg cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacccatcc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgtcactg gcaagatcaa gcagatctg
1201 cgcacatgtggc agggcggtgg ccaggccacc tacggccccc ccatgcgg caacatcacc
1261 tgccgcagca acatcacccg catccctgtc accccgcacg gccgttcaa caccaccaac
1321 aacaccgaga cttccgcggc cggcgccggc gacatgcgcg acaactgcgc cagcggagct
1381 tacaagtaca aggtgggtt gatcaagcccc ctggccatcg ccccccaccaa gccaagcgc
1441 cgcgtgtgc agcgcgagaa ggcgcgcgtg ggcacccgc cccgttctt gggcttctg
1501 ggcgcgcgc gcaagccat gggccgcgc accatccatc tgaccgtca ggcgcgc
1561 ctgcgtgacg gcatctgca gcaagcagac aacctgtga aggccatcga ggcgcgc
1621 cacatgtgc agctgaccgt gtggggcatc aagcagctgc aggccgcgt gctggccatc
1681 gagcgttacc tgaaggacca gcaactgtg ggcacccgtt gctgcacccgg cccgcgttac
1741 tgcaccaccg cctgtccctg gaacagcagc tggagcaaca agagcagaaa ggacatctgg
1801 gacaacatga cctggatgca gtgggaccgc gagatcagca actacacccg cctgatctac
1861 aacccgttgg aggacagccaa gaaccaggcag gagaagaacg agaaggaccc tggagctg
1921 gacaagtggaa acaaccctgtg gaactggcc gacatcagca actggccctg gtacatcaag
1981 atcttcatca tgatctgtgg cggccctgtc ggccctgcac tcatctgc cttgttgc
2041 atcgtgaacc gctgtgcgc gggctacagc cccctgatcc tccagaccctt gacccgc
2101 ccccgccggc tggaccgcctt gggccgcatac gaggaggagg gggccgcacca ggaccgc
2161 cgcacatcc gcctgggtgag cggccctgtt agcctggccctt gggacccatc ggcacaccc
2221 tgcctgttca gctaccaccg cctgcgcgc ttcacccatca tgcgttgc gcccgtgg
2281 ctgcgtggcc acacccgc ggcgcgcgtt cggccgcgtt gggagatcc ttttttttt
2341 ggcgcgcgtt tgcgttgcgtt gggccgtggat ctgaagaaga ggcacccatc cttgttgc
2401 accatgcaca tcaccgtggc cggccgcacc gaccgcacca tcgagctgtt gcaacccatc
2461 tgccgcgc gcaaccat ccccgccgc atccgcagg gcttcgaggc gcccctgt
2521 taactcgat

FIGURE 94

gp160mod.TV1.delV1V2 (SEQ ID NO:122)

1 gaattccatgc gcgtgatggg caccaggaaactgccac gttgggtggat ctggggcata
61 ctggctcttggatgtatgatctgcaaccggaggacc tggggtgac cggtactac
121 ggctgtcccg tggcgccga cgccaaagacc accctgtct gcggcagcga cgccaaaggcc
181 tacgagaccc aggtgcacaa cgtgtggcc accccacgcgcgtgcccac cgacccaaac
241 cccaggaga tcgtgtgg caacgtgacc gagaacitca acatgtggaa gaacgacatg
301 gcccggcaga tgcacggaga cgtgtacgc cgtgtggacc agagcgtgaa gccctgcgtg
361 aagctgaccc ccctgtcggt gggcgccgc aactgcaaca ccagcaccat caccaggcc
421 tgccccaaagg tgagcttcga cccatcccc atccactact ggcggccgc cggtacgc
481 atcctgaagt gcaacaacaa gacctcaac ggcacccggcc cctgtaccaa cgtgagcacc
541 gtgcagtgc cccacggcat caagccgtg gtgagcaccc agctgtctt gaaacggcagc
601 ctggccgagg agggcatcat catccgcgc gagaacctga ccgagaacac caagaccatc
661 atcgtgcacc tgaacgagag cgtggagatc aactgcaacc gcccaacaa caacacccgc
721 aagagcgtgc gcatcgccccc cggccaggcc ttctacgcca ccaacgacgt gatggcaac
781 atccgcccagg cccactgca catcagcacc gaccgtgga acaagaccct gcagcagggt
841 atgaagaagc tggcgagca ctcccaac aagaccatcc agtcaagcc ccacggccgc
901 ggcgacctgg agatcaccat gcacagcttc aactggcg gcgagttctt ctactgcaac
961 accagcaacc tggtaacag cacctaccac agcaacaacg gcacccatcaa gtacaacggc
1021 aacagcagca gccccatcac cctgcagtgc aagatcaagc agatgtgcg catgtggcag
1081 ggcgtggcc aggccaccta cggccccc atgcggcga acatccatcg ccgcagcaac
1141 atcaccggca tccgtctgac cccgcacggc ggctcaaca ccaccaacaa caccgagacc
1201 ttccgccccg gcccggcga catgcgcac aactggcgca gcgagctgtca aagttacaag
1261 gtgggtggaga tcaagccctt gggcatcgcc cccaccaagg ccaagccgc cgtggcag
1321 cggcagaagc gcccgtggg catcgccgc gtttcctgg gtttcctgg cggccggc
1381 agcaccatgg gcccggcag catcaccctg accgtgcagg cccgcccagct gctgagccgc
1441 atcgtgcagc agcagagcaa cctgtcgaa gccatcgagg cccagcagca catgtgcag
1501 ctgaccgtgt gggcatcaa gcagctcgag ccccgctgc tggccatcg ggcgtaccctg
1561 aaggaccaggc agctgtggg catctgggc tgcaaggccgc gctgatctg caccaccgc
1621 gtccctggaa acagcagctg gagcaacaag agcgagaagg acatctgggaa caatcgacc
1681 tggatgcagt gggccggca gatcagcaac tacaccggcc tggatctacaa cctgtggag
1741 gacagccaga accaggcaggaa gaaacgag aaggaccctgc tggagctgaa caatgtggaa
1801 aaccgtgttca actggatcgatc catcacaac tggccctggt acatcaagat cttcatcg
1861 atcgtggcg gctgtatcggt cctgcgcac atcttcggc tggatcgatc gtcgtaccgc
1921 gtgcggcagg gctacaggccc cctgagcttc cagaccctga ccccgccatcg
1981 gaccggctgg gcccgtcgaa ggaggaggc gggcggcagg accggcaccgc cggccatcg
2041 ctggtagcg gcttccgtgg cctggccctgg gacgaccctgc gcaaccctgtg cttgtccagc
2101 taccaccggc tgcgcactt catcctgtatc gccgtgcgcg cctgtggatc gtcgtggcc
2161 agcagcctgc gggccgtcgaa gggccggcgg gagatctgaa agtacctggg cggccatcg
2221 cgtactggg gcttggatc gaaagagc gccatcagcc tggatcgatc gtcgtggcc
2281 accgtggcccg agggcaccga cccgtatcaccatc gagctgtgc gtcgtaccgc
2341 ctgaaatcc cccggccatcgccaggcc ttggatcgatc gtcgtggcc

FIGURE 93

gp140mod.TV1.mut7.delV2 (SEQ ID NO:121)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtgat ctggggcatc
61 ctgggcctct ggatgtgtat gatctcaac accgaggacc tgggggtgac cgtgtactac
121 ggcgtccccg tggcgccgca cgccaagacc accctgtct gcgccagcgca cgccaaggcc
181 tacgagaccc aggtgcacaa cgtgtggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgtggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgacgagga cgtgatcgc ctgtggacc agagcgtgaa gccctgcgtg
361 aagtgcaccc ccctgtgcgt gaccctgaac tgccaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccc gcaacacgac caacaacacc aacggcaccc gcatctacaa catcgaggag
481 atgaagaact gcagctcaa cgccggcgcc ggccgcctga tcaactgaa caccaggacc
541 atcacccagg cctgccccaa ggtgagcttc gacccctatcc ccatccacta ctgcgcctcc
601 gccggctacg ccacccctgaa gtcaacaac aagacccatca acggcacccgg cccctgtac
661 aacgtgagca cctgtgcgtg caccacggc atcaagcccg tggtgagcac ccagctgtg
721 ctgaacggca gcctggccgaa ggagggcata atcatccgca gcgagaacct gaccgagaac
781 accaagacca tcatgtgca cctgaacggag agcgtggaga tcaactgcac cggccccaac
841 aacaacaccc gcaagagcgt gcgcacccggc cccggccagg ccttctacgc caccacgac
901 gtgatggca acatccgcca ggccactgc aacatcagca ccgaccgctg gaacaagacc
961 ctgcagcagg tgaatgaaagaa gctggcgag cacttccca acaagaccat ccagttcaag
1021 ccccacccgg gggcgaccc ggagatcacc atgcacagct tcaactgccc cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacacgcg cagccccatc accctgtcgt gcaagatcaa gcagatgtg
1201 cgcacgtggc agggcggtgg ccaggccacc taaccccccc ccatgcggg caacatcacc
1261 tgccgcagca acatcaccgg catccctgtg acccgccgacg gggcttcaaa caccaccaac
1321 aacaccgaga cttccgccc cggcgccggc gacatgcgcg acaactggcg cagcggcgt
1381 tacaagtaca aggtggtggaa gatcaagcccc ctggccatcg ccccccaccaa ggccatcagc
1441 agcgfgggtgc agagcgagaa gagcgccgtg ggcacccggc cctgtttctt gggcttctg
1501 ggcgcgcgg gcagcaccat ggccgcgcgc agcatcaccc tgaccgtgca ggcccgcag
1561 ctgctgagcgc gcatctgtca gcagcagacg aacctgtga aggccatcga ggcccagcag
1621 cacatgtgc agctgaccgt gtggggcatc aagcagctgc aggccccgt gctggccatc
1681 gagcgctacc tgaaggacca gcagctgtg ggcacccggc gctgcagcgg ccgcctgtac
1741 tgcaccaccgg cctgtccctg gaacacgcg tggagcaaca agagcgagaa ggacatctgg
1801 gacaacatgaa cctggatgca gtggggaccgc gagatcagca actacacccgg ctgtatctac
1861 aacctgtgg aggacagcca gaaccacgcg gagaagaacg agaaggaccg gctggagctg
1921 gacaagtggaa acaacccgtg gaactgggtc gacatcagca actggccctg gtacatctaa
1981 ctcgag

FIGURE 92

gp140mod.TV1.delV2 (SEQ ID NO:120)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtgat ctggggcatc
61 ctgggcttct ggatgtgtat gatctgcaac accgaggacc tgggggtgac cgtgtactac
121 ggcgtccccg tgtggcgca cgccaagacc accctgttct gogccagcga cgccaaggcc
181 tacgagaccc aggtgcacaa cgtgtggcc acccacgcct gogtgcggcc acgacccaaac
241 ccccaggaga tcgtgtggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgacgagga cgtgatcagc ctgtggacc agagctgaa gcccctgogtg
361 aagctgaccc ccctgtgcgt gaccctgaaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccc gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcaacttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc
541 atcacccagg cctggcccaa ggtgagctc gaccctatcc ccatccacta ctgcgcggcc
601 gccggctacg ccatactgaa gtgcaacaac aagacattca acggcaccgg cccctgtac
661 aacgtgagca ccgtgcagtgc accccacgcg atcaagcccg tggtgagcac ccagctgtg
721 ctgaaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac
781 accaagacca tcatctgtca cctgaacgag agcgtggaga tcaactgac cccggccaaac
841 aacaacaccc gcaagagcgt ggcacatcgcc cccggccagg cttctacgc caccaacgc
901 gtgatcgca acatccgca ggccactgca aacatcagca cgcaccgtg gaacaagacc
961 ctgcagcagg ttagtgaagaa gctggcgag cacttccca acaagaccat ccagttcaag
1021 cccacccgg gcccggaccc ggagatcacc atgcacagct tcaactgcgg cggcgagtgc
1081 ttctactgca acaccagcaa cctgttcaac agcaccta acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatgtg
1201 cgcacatgtggc agggcgtggg ccaggccacc taacggccccc ccatcgccgg caacatcacc
1261 tgccgcagca acatcaccgg catcctgtgc acccgcgacg gggcttcaa caccaccaac
1321 aacaccgaga ccttccggcc cggccggccg gacatgcgcg acaactggcg cagcgagctg
1381 tacaagtaca aggtggatgg gatcaagccc ctgggcacatcg ccccccaccaa ggccaagcgc
1441 cgcgtggtc agcgcgagaa ggcgcggcg ggcacatcgcc cctgttctc gggcttctg
1501 ggccggccgg gcagcaccat gggccggcc accatcaccc tgacgtgca ggccggccag
1561 ctgctgagcg gcatctgtca gacgcacggc aacctgtga aggccatcga ggcccagcag
1621 cacatgtgc agctgaccgt gtggggcatc aagcagctgc aggcccccggt gctggccatc
1681 gagcgcgtacc tgaaggacca gcatctgtgg gtcgcagcgg cccctgtac
1741 tgcaccaccc cctgtccctg gaacagcagc tggagcaaca agagcgagaa ggacatctgg
1801 gacaacatga cctggatgca gtggggaccgc gagatcagca actacaccgg cctgatctac
1861 aacctgtgg aggacagcca gaaccagcag gagaagaacg agaaggaccc gctggagctg
1921 gacaagtggaa acaacatgtg gactgggttc gacatcagca actggccctg gtacatctaa
1981 ctcgag

FIGURE 91

gp120mod.TV1.delV2 (SEQ ID NO:119)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtgat ctggggcatc
61 ctggcttc ggtatgcgtat gatctgcac accgaggacc tggggtgac cgtgtactac
121 ggcgtgccc tggcgccga cgccaagacc accctttct ggcgcgcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtggcc acccacgcct gcgtgcccac cgaccccaac
241 cccccaggaga tcgtgtggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgacgagga cgtgatcgc ctgtggacc agagctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctaca catcgaggag
481 atgaagaact gcagcttcaa cgccggcgcg ggcgcctga tcaactgaa caccagcacc
541 atcacccagg cctgccccaa ggtgagctt gacccatcc ccatccacta ctgcgcccc
601 gccggctacg ccatcctgaa gtgcaacaac aagaccttca acggcaccgg cccctgctac
661 aacgtgagca cctgtcagt caccacggc atcaagcccg tggtgagcac ccagctgtg
721 ctgaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac
781 accaagacca tcatcgtgca cctgaacggc agcgtggaga tcaactgac ccgcccccaac
841 aacaacaccc gcaagagcgt ggcacatggc cccggccagg cttctacgc caccacgac
901 gtgatcggca acatccgcca ggccactgc aacatcagca cgcaccgtg gaacaagacc
961 ctgcagcagg ttagtgaagaa gctggcgag cacttccca acaagaccat ccagtcaag
1021 cccccacggc gggcgaccc ggagatcacc atgcacagct tcaactgccc cggcgagtc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatgt
1201 cgcacatgtggc agggcggtt ccaggccacc tacggccccc ccatcgccgg caacatcacc
1261 tgccgcagca acatcaccgg catctgtgt acccgcgacg gggcttcaa caccaccaac
1321 aacaccgaga ccttccccc cggggcgcc gacatgcgcg acaactggcg cagcgagctg
1381 tacaagtaca aggtggtgg aatcaagccc ctgggcattt ccccccaccaa ggcacagcgc
1441 cgcgtgtgc agcgcgagaa ggcgttaactc gag

FIGURE 90

Vpu_TV2_C_ZAwt (SEQ ID NO:118)

ATGTTAGATTAACTGCAAGAACATAGATTCTAGATTAGGAATAGGAGCATTGA
TAGTAGCACTAACATAGCAATAATAGTGTGGACCATAGTATATAGAATA
TAGGAAATTGGTAAGGCAAAGGAAAATAGACTGGTTAGTAAAAGGATTAG
GGAAAGAGCAGAACAGACAGTGGCAATGAGAGCGAGGGGGATACTGAAGAATT
ATCGACACTGGTGGATATGGGGCATCTTAGGCTTGATGCTAATGATGTGT
AA

FIGURE 89

Vpu_TV2_C_ZAopt (SEQ ID NO:117)

ATGCTGGACCTGACCGCCCGCATCGACAGCCGCCTGGGCATCGGCCCTGA
TCGTGGCCCTGATCATGCCATCATCGTGTGGACCATCGTGTACATCGAGTAC
CGCAAGCTGGTGCGCCAGCGCAAGATCGACTGGCTGGTGAAGCGCATCCGCG
AGCGCGCCGAGGGACAGCGCAACGAGAGCGAGGGCGACACCGAGGAGCTGA
GCACCCTGGTGGACATGGGCCACCTGCGCCTGCTGGACGCCAACGACGTGTA
A

FIGURE 88

Vpr_TV2_C_ZAwt (SEQ ID NO:116)

ATGGAACAAGCCCCAGAACGACCAGGGGCCAGAGGGAACCATACAATGAA
TGGACACTAGAGCTTTAGAAGAACTCAAGCAGGAAGCTGTAGACACTTC
CTAGACCATGGCTCCATAACTTAGGACAACATATCTATGAAACCTATGGAGA
TACTTGGACAGGAGTTGAAGCAATAATAAGAATCCTGCAACAATTACTGTTT
ATTCATTCAGGATTGGGTGCCATCATAGCAGAATAGGCATTTGCGACAGA
GAAGAGCAAGAAATGGAGCCAATAGATCC

FIGURE 87

Vpr_TV2_C_ZAopt (SEQ ID NO:115)

ATGGAGCAGGCCCGAGGACCAGGGCCCCAGCGCGAGCCCTACAAACGAG
TGGACCCCTGGAGCTGCTGGAGGAGCTGAAGCAGGAGGCCGTGCGCCACTTCC
CCCGCCCCCTGGCTGCACAACCTGGGCCAGCACATCTACGAGACCTACGGCGA
CACCTGGACCGGCGTGGAGGCCATCATCCGCATCCTGCAGCAGCTGCTGTTTC
ATCCACTTCCGCATCGGCTGCCACCACAGCCGCATGGCATCCTGCGCCAGC
GCCGCGCCCGCAACGGCGCCAACCGCAGC

FIGURE 86

Vif_TV2_C_ZAwt (SEQ ID NO:114)

ATGGAAAACAGATGGCAGGTGCTGATTGTGTGGCAGGTAGACAGGATGAAG
ATTAGAACATGGCACAGTTAGTAAAGCACCATAATGTATGTTCGAGGAGAG
CTGATGGATGGTTCTACAGACATCATTATGAAAGCAGACACCCAAAAGTAAG
TTCAGAAGTACACATCCCCATTAGGAGATGCCAGGTTAGTAATAAAAACATAT
TGGGGTCTGCAGACAGGAGAAAGAGCTTGGCATTGGGTACGGAGTCTCCA
TAGAATGGAGATTGAGAAGATATAGCACACAAGTAGACCCCTGACCTGACAG
ACCAACTAATTATGCATTATTTGATTGTTGCAGAATCTGCCATAAGG-
AAAGCCATACTAGGACAGATAGTTAGCCCTAAGTGTGACTATCAAGCAGGAC
ATAACAAGGTAGGATCTCTACAATACTTGGCACTGACAGCATTGATAAAACC
AAAAAAAGATAAAGCCACCTCTGCCTAGTGTAGGAAATTAGTAGAGGGATAGA
TGGAAACAAGCCCCAGAAGACCAGGGGCCAGAGGGAACCATAATGAAT
GGACACTAG

FIGURE 85

Vif_TV2_C_ZAopt (SEQ ID NO:113)

ATGGAGAACCGCTGGCAGGTGCTGATCGTGTGGCAGGTGGACCGCATGAAGA
TCCGCACCTGGCACAGCCTGGTGAAGCACCACATGTACGTGAGCCGCCGC
CGACGGCTGGTTCTACCGCCACCAACTACGAGAGCCGCCACCCCAAGGTGAGC
AGCGAGGTGCACATCCCCCTGGCGACGCCGCCCTGGTGATCAAGACCTACT
GGGGCCTGCAGACCGCGAGCGCGCCTGGCACCTGGCCACGGCGTGAGCA
TCGAGTGGCGCCTGCGCCGCTACAGCACCCAGGTGGACCCGACCTGACCGA
CCAGCTGATCCACATGCACTACTCGACTGCTTCGCCAGAGAGCGCCATCCGC
AAGGCCATCCTGGGCCAGATCGTAGGCCCAAGTGCAGTACCGGCCGGCC
ACAACAAGGTGGCAGCCTGCAGTACCTGGCCCTGACCGCCCTGATCAAGCC
CAAGAAGATCAAGCCCCCTGCCAGCGTGCAGCAAGCTGGTGGAGGACCGC
TGGAAACAAGCCCCAGAAGACCCGGCCGCCGGCAACCACACCATGAAC
GGCCACTAG

FIGURE 84

TatExon2_TV2_C_ZAwt (SEQ ID NO:112)

CCCTTATCCCAAACCGAGGGGACCCGACAGGCTCGGAGGAATCGAAGAAG
AAGGTGGAGAGCAAGACAGCAGCAGATCCATTGATTAG

FIGURE 83

TatExon2_TV2_C_ZAopt (SEQ ID NO:111)

CCCTGAGCCAGACCGCGGGCAGCCCCACCGGCAGCGAGGAGAGCAAGAAG
AAGGTGGAGAGCAAGACCGCCGCCGACCCCTTCGACTAG

FIGURE 82

TatExon1_TV2_C_ZAwt (SEQ ID NO:110)

ATGGAGCCAATAGATCTAACCTAGAACCCCTGGAACCATCCAGGAAGTCAGC
CTAAAACGTGTTGTAATGGGTGTTACTGTAAACGTTGCAGCTATCATTGTCTA
GTTTGCTTCAGAAAAAAAGGCTTAGGCATTACTATGGCAGGAAGAAGCGGA
GACAGCGACGAAGCGCTCCTCCAAGCAATAAAGATCATCAAGATCCTCTACC
AAAGCAG

FIGURE 81

TatExon1_TV2_C_ZAopt (SEQ ID NO:109)

ATGGAGCCCATCGACCCCCAACCTGGAGCCCTGGAACCACCCCGGCAGCCAGC
CCAAGACCGCCTGCAACGGCTGCTACTGCAAGCGCTGCAGCTACCACTGCCT
GGTGTGCTTCCAGAAGAAGGGCCTGGCATCTACTACGCCGCAAGAACGCG
GCCAGCGCCGAGGCCCGCCAGCAACAAGGACCACCAGGACCCCTGC
CCAAGCAG

FIGURE 80

RevExon2_TV2_C_ZAwt (SEQ ID NO:108)

ACCCCTTATCCCAAACCGAGGGGACCCGACAGGCTCGGAGGAATCGAAGAA
GAAGGTGGAGAGCAAGACAGCAGCAGATCCATTGATTAGTGAGCGGATTCT
TGACACTTGCCTGGGACGACCTACGAAGCCTGTGCCCTTGCTACCACCGA
TTGAGAGACTTCATATTAATTGTAGTGAGAGCAGTGGAACTTCTGGGACACA
GTAGTCTCAGGGGACTGCAGAGGGGTGGGAACCCCTAA

FIGURE 79

RevExon2_TV2_C_ZAopt (SEQ ID NO:107)

CCCTACCCCAAGCCCCGAGGGCACCCGCCAGGCCGCCGCAACCGCCGCC
GCTGGCGCGCCGCCAGCAGCAGATCCACAGCATCAGCGAGCGCATCCTGGA
CACCTGCCTGGGCCGCCCAACCAAGCCGTGCCCTGCTGCTGCCCTCATCG
AGCGCCTGCACATCAACTGCAGCGAGAGCAGCGGCCACCAGCGGCACCCAGT
AGAGCCAGGGCACCGCCGAGGGCGTGGCAACCCCTAA

FIGURE 78